Kongeriget Danmark

Patent application No.:	0512/97
Date of filing:	05 May 1997
Applicant:	Dako A/S, att. Mariannne H. Christensen, Produktionsvej 42, 2600 Glostrup, DK

This is to certify the correctness of the following information:

The attached photocopy is a true copy of the following document:

- The specification, claims, abstract and drawings as filed with the application on the filing date indicated above.

Erhvervsministeriet
Patentdirektoratet

TAASTRUP 08 Aug 1997

Jytte Hansen

Jytte Hansen
Kontorfuldmægtig

0512/97 05 MAJ 97

NOVEL PROBES FOR THE DETECTION OF MYCOBACTERIA

The present invention relates to novel probes and to mixtures of such probes, in addition to the design, construction and use of such novel probes or a mixture thereof for detecting the presence of mycobacteria, which probes are capable of detecting the organisms in test samples, e.g. sputum, expectorates, aspirates, cerebrospinal fluid, urine, blood and tissue sections, food, soil and water. The invention relates in particular to novel probes and mixtures thereof for the detection of mycobacteria of the Mycobacterium tuberculosis Complex (MTC) and the detection of one or more mycobacteria other than mycobacteria of the Mycobacterium tuberculosis Complex. The invention further relates to diagnostic kits comprising one or more of such probes.

BACKGROUND OF THE INVENTION

Mycobacteria are slow growing, acid fast, aerobic bacilli. At least nineteen Mycobacterium species have been associated with disease in humans, among them M. tuberculosis, M. bovis, M. avium, M. intracellulare and M. leprae. Species that are not normally pathogenic to healthy individuals may cause disease in immunocompromised individuals, e.g. those infected with HIV. M. kansasii and M. xenopi cause lung infection, and infection with M. marinum causes arthritis and osteomyelitis. Other clinically relevant mycobacteria are M. microti, M. paratuberculosis, M. scrofulaceum, M. africanum M. gordonea, M. chelonei and M. fortuitum. The MTC group includes M. tuberculosis, M. bovis, M. kansasii and M. africanum. M. avium, M. intracellulare, M. paratuberculosis and M. lepraemurium are included in a group named Mycobacterium avium Complex. Classification and further description of the various mycobacteria can be found in e.g. Clinical Microbiology Reviews, 1-25 (January 1992) and Clinical Microbiology Reviews, 266-310 (July 1993).

One very life-threatening and highly epidemic disease is tuberculosis caused by infection with mycobacteria of the MTC group, in particular M. tuberculosis. Tuberculosis is presently the predominant infectious cause of morbidity and mortality world-wide, and is estimated to kill about three million people annually. WHO estimates that the annual number of new cases of tuberculosis will increase from 7.5 million in 1990 to 10.2 million in 2000, an escalation that will result in approximately 90 million new cases during this decade. It is furthermore estimated that 30 million people will die from tuberculosis during the 1990s, which equals one quarter of preventable deaths among adults.

The prevalence of tuberculosis has been very high in the poorer parts of the world such as Asia, Africa and South-America, but in recent years an increase has also been observed in

industrialised countries. This appears to be due to an interaction of various factors including i.a. patterns of migration, poorly organised tuberculosis programmes and nutrition problems. Furthermore, a serious threat will arise from the emergence of new strains that are multi-drug resistant.

Considering the perspective and impact the disease has, the development of rapid, specific and preferably easy-performed and economic feasible diagnostic detection tests are of utmost importance and would be a very valuable tool in the fight against the spread of tuberculosis.

Presently, the detection of mycobacteria by microscopy gives the more accurate diagnosis. The sample (e.g. an expectorate) is stained for the presence of acid-fast bacilli using Ziehl-Neelsen staining. However, Ziehl-Neelsen staining does not provide the necessary information about the type of infection, only whether acid fast bacilli are present in the sample. Ziehl-Neelsen positive samples may subsequently be cultured. Cultivation is sensitive, and it may be possible to detect 10-100 organisms per sample, but the result is not available before up to 8 weeks of cultivation. Likewise, information of drug susceptibility is not available until after 1-3 weeks of further testing.

Since Ziehl-Neelsen staining cannot be used to determine whether the infection is caused by mycobacteria of the MTC group or mycobacteria other than mycobacteria of the MTC group, a positive staining leads frequently to very costly isolation of all patients as well as treatment with medicaments to which the patient may not even respond. Species identification is presently carried out following cultivation using traditional biochemical methods or probe hybridisation assays (e.g. AccuProbe by Gen-Probe Inc.). There is, therefore, an increasing need for means allowing a more rapid distinction between mycobacteria of the MTC group and mycobacteria other than those of the MTC group.

Automated detection is rapidly becoming available for large scale testing for the presence of mycobacteria. Such systems include ESP Myco Culture System (Difco), MB/BacT (Organon Teknika) and MGIT (Becton Dickinson). These test methods are based on colorimetric or fluorometric detection of carbon dioxide or oxygen produced by mycobacterial metabolism.

Some of the attempts to replace the methods based on cultivation rely on target amplification. One of such newly developed target amplification method is based on PCR. The principle of this reaction is, through amplification of specific nucleic acid sequences of the mycobacteria, to increase the copy number of the specific sequence to a level where it may be detectable in an early stage of the infection. In principle, the PCR reaction offers the possibility of detecting as few as one target sequence. In most cases, the DNA is extracted prior to carrying out the

PCR reaction. However, it has become clear that the target amplification test cannot replace culture test as only samples which are Ziehl-Neelsen positive give a satisfactory sensitivity.

Furthermore, false negative results may be obtained due to the presence of inhibitors of the PCR reaction such as haemoglobin and other proteins.

Another problem arises from cross-contamination of negative specimens with bacteria not present in the sample. This may cause problems in conventional bacteriological procedures and may lead to a positive PCR result. Contamination of reagents and specimens with amplified PCR products is yet another well-recognised problem when using a PCR-based test.

Nucleic acid probes for detecting rRNA of mycobacteria have been described in for example US 5 547 842, EP-A 0 572 120 and US 5 422 242.

SUMMARY OF THE INVENTION

The present invention relates to novel peptide nucleic acid probes and to mixtures of such probes for detecting the presence of mycobacteria in a sample. In accordance with claim 1, the probes are directed to sequences in rRNA and genomic sequences corresponding to said rRNA (rDNA). rRNA is present in a high number of copies in each cell, and hence a well suited target for a sensitive test. The probes are, as defined in claim 2, suitably directed to target sequences in the 23S, 16S or 5S rRNA or DNA coding for said rRNA.

Thus, in a first aspect, the invention features a hybridisation assay probe and a mixture of such probes for detecting the presence of mycobacteria in accordance with claim 1 and 2. Such probes should not to any significant degree cross react with nucleic acid from other organisms in the test sample under appropriate stringency conditions. Cross reactivity to organisms that are unlikely to appear in the sample may not be of importance. In in situ assays implying examination by microscopy, it is possible to distinguish mycobacteria from other bacteria based on evaluation of the morphology of the observed bacilli.

In another aspect, the invention relates to novel peptide nucleic acid probes for detecting the presence of mycobacteria of the MTC group, in particular M. tuberculosis, and one or more mycobacteria other than mycobacteria of the MTC group, in particular mycobacteria of the M. avium Complex (claim 3).

Claims 4 to 6 and 9 to 14 relate to probes for detecting the presence of mycobacteria of the MTC group. Claims 7 to 13 relate to probes for detecting one or more mycobacteria other than

mycobacteria of the MTC group. Claims 15 relates to a mixture of peptide nucleic acid probes according to claims 1 to 14.

In a further aspect, as defined in claims 16 to 18, the invention relates to the use of the peptide nucleic acid probes and mixtures of such probes according to claims 1 to 15.

The present invention also relates to a method for detecting the presence of mycobacteria in accordance with claims 19 to 26.

In yet another aspect, the present invention relates to a kit (claim 27 and 28) comprising at least one peptide nucleic acid probe as defined in claims 1 to 14.

BRIEF DESCRIPTION OF THE FIGURES

Alignments of rRNA sequences of M. tuberculosis (as a representative of the MTC group) and important closely related species thereto, and M. avium (as a representative of the mycobacterial other than those of the MTC group) and important closely related species thereto for the 23S, 16S and 5S rRNA have been made (Figures 1A-1K, 2A-2D, 3, 4A-4N and 5A-B).

Alignment for the MTC group (23S rRNA)

Figures 1A-1K show alignments of 23S rDNA sequences of M. tuberculosis (GenBank entry GB:MTCY130, accession number Z73992), M. avium (GenBank entry GB:MA23SRNA, accession number X74494), M. paratuberculosis (GenBank entry GB:MPARRNA, accession number X74495), M. phlei (GenBank entry GB:MP23SRNA, accession number X74493), M. leprae (GenBank entry GB:ML5S23S, accession number X56657), M. gastri (GenBank entry GB:MG23SRRNA, accession number Z17211), M. kansasii (GenBank entry GB:MK23SRRNA, accession number Z17212), and M. smegmatis (GB:MS16S23S5, accession number Y08453). Preferred peptide nucleic acid probes should enclose at least one nucleobase complementary to a nucleobase of M. tuberculosis 23S rRNA within positions 149-158, 220-221, 328-361, 453-455, 490-501, 637-660, 706-712, 762-789, 989, 1068-1072, 1148, 1311-1329, 1361-1364, 1418, 1563-1570, 1627-1638, 1675-1677, 1718, 1734-1740, 1967-1976, 2403-2420, 2457-2488, 2952-2956, 2966-2969, 3000-3003, and 3097-3106 of the alignment (indicated by heavy frames). Mismatches between the sequences of M. avium, M. phlei, M. leprae, M. paratuberculosis, M. gastri and M. kansasii and that of M. tuberculosis in the alignment are indicated by light frames.

Alignment for the MTC group (16S rRNA)

Figures 2A-2D show alignments of 16S rDNA sequences of M. tuberculosis (GenBank entry GB:MTU16SRN, accession number X52917), M. bovis (GenBank entry GB:MSGTGDA, accession number M20940), M. avium (GenBank entry GB:MSGRRDA, accession number M61673), M. intracellulare (GenBank entry GB:MIN16SRN, accession number X52927), M. paratuberculosis (GenBank entry GB:MSGRRDH, accession number M5680), M. scrofulaceum (GenBank entry GB:MSC16SRN, accession number X52924), M. leprae (GenBank entry GB:MLEP16S1, accession number X55587), M. kansasii (GenBank entry GB:MKRRN16, accession number X15916), and M. gastri (GenBank entry GB:MGA16SRN, accession number X52919). Preferred peptide nucleic acid probes should enclose at least one nucleobase complementary to a nucleobase of M. tuberculosis 16S rRNA within positions 76-79, 98-101, 135-136, 194-201, 220-229, 242, 474, 1136-1145, 1271-1272, 1287-1292, 1313, and 1334 of the alignment (indicated by heavy frames). Mismatches between the sequences of M. bovis, M. avium, M. intracellulare, M. paratuberculosis, M. scrofulaceum, M. lepeae, M. kansasii, and M. gastri and that of M. tuberculosis in the alignment are indicated by light frames.

Alignment for the MTC group (5S rRNA)

Figure 3 shows alignments of 5S rDNA sequences of M. tuberculosis (GenBank entry GB:MTDNA16S, accession number x75601), M. bovis (GenBank entry GB:MBRRN5S, accession number X05526), M. phlei (GenBank entry GB:MP5SRRNA, accession number X55259), M. leprae (GenBank entry GB:ML5S23S, accession number X56657), and M. smegmatis (GenBank entry GB:MS16S23, accession number Y08453). Preferred peptide nucleic acid probes should enclose at least one nucleobase complementary to a nucleobase of M. tuberculosis 5S rRNA within position 86-90 of the alignment (indicated by heavy frame). Mismatches between the sequences of M. bovis, M. phlei, M. leprae, M. smegmatis and M. luteus and that of M. tuberculosis in the alignment are indicated by light frames.

Alignment for Mycobacteria other than those of the MTC group (23S)

Figures 4A-4N show alignments of 23S rDNA sequences of M. avium (GenBank entry GB:MA23SRNA, accession number X74494), M. paratuberculosis (GenBank entry GB:MPARRNA, accession number X74495), M. tuberculosis (GenBank entry GB:MTY130, accession number Z73992), M. phlei (GenBank entry GB:MP23SRNA, accession number X74493), M. leprae (GenBank entry GB:ML5S23S, accession number X56657), M. gastri (GenBank entry GB:MG23SRRNA, accession number Z17211), M. kansasii (GenBank entry GB:MK23SRRNA, accession number Z17212), and M. smegmatis (GB:MS16S23S5, accession number Y08453). Preferred peptide nucleic acid probes should enclose at least one nucleobase complementary to a nucleobase of M. avium 23S rRNA within positions 99-101, 183, 261-271, 281-284, 290-293, 327-335, 343-357, 400-405, 453-462, 587-599, 637-660,

704-712, 763-789, 1060-1074, 1177-1185, 1259-1265, 1311-1327, 1345-1348, 1556-1570, 1608-1613, 1626-1638, 1651-1659, 1675-1677, 1734-1741, 1847-1853, 1967-1976, 2006-2010, 2025-2027, 2131-2232, 2252-2255, 2396-2405, 2416-2420, 2474-2478, 2687, 2719, 2809, 3062-3068, and 3097-3106 of the alignment (indicated by heavy frames). Mismatches between the sequences of M. paratuberculosis, M. tuberculosis, M. phlei, M. leprae, M. gastri, M. kansasii, M. luteus, and M. smegmatis and that of M. avium in the alignment are indicated by light frames.

Alignment for Mycobacteria other than those of the MTC group (16S)

Figures 5A-5B show alignments of 16S rDNA sequences of M. avium (GenBank entry GB:MSGRRDA, accession number X52917), M. intracellulare (GenBank entry GB:MIN16SRN, accession number X52927),M. paratuberculosis (GenBank entry GB:MSGRRDH, accession number M5680), M. scrofulaceum (GenBank entry GB: MSC16SRN, accession number X52924), M. tuberculosis (GenBank entry GB:MTU16SRN, accession number X52917), M. bovis (GenBank entry GB:MSGTGDA, accession number M20940), M. leprae (GenBank entry GB:MLEP16S1, accession number X55587), M. kansasii (GenBank entry GB:MKRRN16, accession number X159916), and M. gastri (GenBank entry GB:MGA16SRN, accession number X52919). Preferred peptide nucleic acid probes should enclose at least one nucleobase complementary to a nucleobase of M. avium 16S rRNA within positions 135-136, 472-475, 1136-1144, 1287-1292, 1313, and 1334 of the alignment (indicated by heavy frames). Mismatches between the sequences of M. intracellulare, M. paratuberculosis, M. scrofulaceum, M. tuberculosis, M. bovis, M. leprae, M. kansasii, and M. gastri and that of M. avium in the alignment are indicated by light frames.

SPECIFIC DESCRIPTION

The present invention provides novel probes for use in rapid and sensitive hybridisation based assays for the detection of mycobacteria, in particular mycobacteria of the MTC group, and one or more mycobacteria other than mycobacteria of the MTC group.

We have identified suitable variable regions of the target nucleic acid by comparative analysis of generally available rRNA sequences. Computers and computer programs, which have been used for the purposes disclosed herein, are generally available. From such alignments, possibly suitable probes can be identified. The alignments are thus a useful guideline for designing probes with desired characteristics. The extent and specificity of hybridisation between the probe and its target are affected by a number of factors, whereby manipulation of one or more of those factors will determine the exact sensitivity and specificity of a probe in question.

When designing the probes, due regard should be taken to the assay conditions under which the probes are to be used. The stringency of the assay conditions determines the degree of complementarity needed between the probe and nucleic acid for formation of hybrids. Stringency is chosen so as to maximise the difference in stability between the hybrid formed with the target nucleic acid and that formed with the non-target nucleic acid. It will typically be necessary to choose high stringency conditions for probes which specificity depend on only one mismatch to non-target sequences. The more mismatches, the less demand for high stringency conditions.

Furthermore, probes should be positioned so as to minimise the stability of probe:non-target nucleic acid hybrids. This may be accomplished by minimising the degree of complementarity to non-target nucleic acid, i.e. by designing the probe to span as many destabilising mismatches as possible, and to include as many additions/deletions relative to the target sequence as possible. Whether a probe is useful to detect a particular mycobacterial species depends largely on the thermal stability difference between probe:target hybrids and probe:non-target hybrids. For rRNA targets, however, the secondary structure of the region of the rRNA molecule in which the target sequence is located may also be of importance.

Hybrids formed between peptide nucleic acid probes and nucleic acids have a higher thermal instability of mismatching bases compared to nucleic acid duplexes of the same sequences. Thus, the peptide nucleic acid probes exhibit a greater specificity for a target nucleic acid sequence than a traditional nucleic acid probe, which is seen as a greater difference in T_m values for probe:target hybrids and probe:non-target hybrids.

The primary concern regarding the length of the peptide nucleic acid probes is the warranted specificity, i.e. what length is specific enough for that particular application. The optimal length of an oligomer probe comprising a particular site with differences in base composition, e.g. among selected regions of mycobacterial rRNA, is a compromise between the general pattern that longer probes ensure specificity and shorter probes ensure that the destabilising differences in base composition constitute a greater portion of the probe.

Peptide nucleic acids can form duplexes in either orientation, but the antiparallel orientation forms the most regular and stable duplex. Hence the antiparallel configuration is preferred for probe applications. Triplex formation with a stoichiometry of two peptide nucleic acid strands and one nucleic acid strand may occur if the peptide nucleic acid has a high pyrimidine content. Such triplexes are very stable, and probes capable of forming triplexes may thus be suitable for certain applications.

Mainly because the peptide nucleic acid strand is uncharged, a peptide nucleic acid-nucleic acid-duplex will have a higher T_m than the corresponding nucleic acid-nucleic acid-duplex. Typically there will be an increase in T_m of about 1 °C per base pair at 100 mM NaCl depending on the sequence (Egholm et al. (1993), Nature, 365, 566-568).

In contrast to DNA-DNA-duplex formation, no salt is necessary to facilitate and stabilise the formation of a peptide nucleic acid-DNA or a peptide nucleic acid-RNA duplex. The T_m of the peptide nucleic acid-DNA-duplex changes only little with increasing ionic strength. Typically for a 15-mer, the T_m will drop only 5 °C when the salt concentration is raised from 10 mM NaCl to 1 M NaCl. At low ionic strength (e.g. 10 mM phosphate buffer with no salt added), hybridisation of a peptide nucleic acid to a target sequence is possible under conditions where no stable DNA-DNA-duplex formation occurs. Furthermore, target sites that normally are inaccessible can be made more readily accessible for hybridisation with peptide nucleic acid probes at low salt concentration as the secondary and tertiary structure of nucleic acids are melted under such conditions.

Although it is preferred to use peptide nucleic acid probes targeting specific sequences of rRNA, it will readily be understood that peptide nucleic acid probes complementary to the rRNA targeting probes will be useful for the detection of the genes (DNA) coding for said sequence specific rRNA (rDNA).

In the broadest aspect, the present invention relates to peptide nucleic acid probes for detecting the presence of mycobacteria in a sample (claims 1 to 3). Peptide nucleic acids are non-naturally occurring polyamides or polythioamides which can bind to nucleic acids (DNA and RNA) with sequence specificity as described e.g. in US 5 539 082.

In accordance with the present invention, peptide nucleic acid probes of formula (I) or a mixture thereof as defined in claims 4 to 8 are provided, which probes are useful for detecting mycobacteria of the MTC group (claims 4 to 6) or of one or more mycobacteria other than mycobacteria of the MTC group in a sample (claims 7 to 8). The probes are directed to 23S, 16S or 5S rRNA, with the provisos indicated in claims 4 to 8.

In the present context and the claims, the term "naturally occurring nucleobases" includes the four main DNA bases (i.e. thymine (T), cytosine (C), adenine (A) and guanine (G)) as well as other naturally occurring nucleobases (e.g. uracil (U) and hypoxanthine).

The term "non-naturally occurring nucleobases" comprises i.a. modified naturally occurring

nucleobases. Such non-naturally occurring nucleobases may be modified by substitution by e.g. one or more C_{1-8} alkyl, C_{1-8} alkenyl or C_{1-8} alkynyl groups or labels. Examples of non-naturally occurring nucleobases are purine, 2,6-diamino purine, 5-propynylcytosine (C propynyl), isocytosine (iso-C), 5-methyl-isocytosine ($iso^{Me}C$) (see e.g. Tetrahedron Letters Vol 36, No 12, 2033-2036 (1995) or Tetrahedron Letters Vol 36, No 21, 3601-3604 (1995)), 7-deazaadenine, 7-deazaguanine, N^4-ethanocytosine, N^6-ethano-2,6-diaminopurine, 5-(C_{3-6})-alkenyluracil, 5-(C_{3-6})-alkynylcytosine, 5-fluorouracil and pseudocytosine.

Examples of useful intercalators are e.g. acridin, antraquinone, psoralen and pyrene.

Examples of useful nucleobase-binding groups are e.g. groups containing cyclic or heterocyclic rings. Non-limiting examples are 3-nitro pyrrole and 5-nitro indole.

It is to be understood that alkyl, alkenyl and alkynyl groups may be branched or non-branched, cyclic or non-cyclic, and may further be interrupted by one or more heteroatoms, or may be unsubtituted or substituted by or may contain one or more functional groups. Non-limiting examples of such functional groups are acetyl groups, acyl groups, amino groups, carbamido groups, carbamoyl groups, carbamyl groups, carbonyl groups, carboxy groups, cyano groups, dithio groups, formyl groups, guanidino groups, halogens, hydrazino groups, hydrazo groups, hydroxamino groups, hydroxy groups, keto groups, mercapto groups, nitro groups, phospho groups, phosphono groups, phospho ester groups, sulfo groups, thiocyanato groups, cyclic, aromatic and heterocyclic groups.

C_{1-4} groups contain from 1 to 4 carbon atoms, C_{1-6} groups contain from 1 to 6 carbon atoms, and C_{1-15} groups contain from 1 to 15 carbon atoms, not including optional substituents, heteroatoms and/or functional groups. Non-limiting examples of such groups are $-CH_3$, $-CF_3$, $-CH_2-$, $-CH_2CH_3$, $-CH_2CH_2-$, $-CH(CH_3)_2$, $-OCH_3$, $-OCH_2-$, $-OCH_2CH_3$, $-OCH_2CH_2-$, $-OCH(CH_3)_2$, $-OC(O)CH_3$, $-OC(O)CH_2-$, $-C(O)H$, $-C(O)-$, $-C(O)CH_3$, $-C(O)OH$, $-C(O)O-$, $-CH_2NH_2$, $-CH_2NH-$, $-CH_2OCH_3$, $-CH_2OCH_2-$, $-CH_2OC(O)OH$, $-CH_2OC(O)O-$, $-CH_2C(O)CH_3$, $-CH_2C(O)CH_2-$, $-C(O)NH_2$, $-CH{=}CH_2$, $-CH{=}CH-$, $-CH{=}CHCH_2C(O)OH$, $-CH{=}CHCH_2C(O)O-$, $-C{\equiv}CH$, $-C{\equiv}C-$, $-CH_2C{\equiv}CH$, $-CH_2C{\equiv}C-$, $-CH_2C{\equiv}CCH_3$, $-OCH_2C{\equiv}CH$, $-OCH_2C{\equiv}C-$, $-OCH_2C{\equiv}CCH_3$, $-NHCH_2C(O)-$, $-NHCH_2CH_2C(O)-$, $-NH(CH_2CH_2O)_2CH_2C(O)-$, and $HO(O)CCH_2C(O)(NH-(CH_2CH_2O)_2CH_2C(O))_2-$, phenyl, benzyl, naphthyl, oxazolyl, pyridinyl, thiadiazolyl, triazolyl, and thienyl.

Within the present context, the expression "naturally occurring amino acid" is intended to comprise D- and L-forms of amino acids commonly found in nature, e.g. D- and L-forms of Ala (alanine), Arg (arginine), Asn (aspargine), Asp (aspartic acid), Cys (cysteine), Gln (glutamine),

Glu (glutamic acid), His (histidine), Ile (isoleucine), Leu (leucine), Lys (lysine), Met (methionine), Phe (phenylalanine), Pro (proline), Ser (serine), Thr (threonine), Trp (tryptophan), Tyr (tyrosine) and Val (valine).

In the present context, the expression "non-naturally occurring amino acid" is intended to comprise D- and L-forms of amino acids other than those commonly found in nature as well as modified naturally occurring amino acids. Examples of useful non-naturally occurring amino acids are D- and L-forms of β-Ala (β-alanine) Cha (cyclohexylalanine), Cit (citrulline), Hci (homocitrulline), HomoCys (homocystein), Hse (homoserine), Nle (norleucine), Nva (norvaline), Orn (ornithine), Sar (sarcosine) and Thi (thienylalanine).

The strength of the binding between the probe and the nucleic acid sequence may further be influenced by the ligand Q. When Q designates a nucleobase, Hoogsteen and/or Watson-Crick base pairing assist(s) in the formation of hybrids between a nucleic acid sequence to be detected and the probe. It is contemplated that one or more of the ligands may be a group which contribute little or none to the binding of the nucleic acid such as hydrogen. It is contemplated that suitable probes to be used comprise less than 25% by weight of peptide nucleic acid moieties, wherein Q designates such groups. One or more of the ligands Q may be groups that stabilise nucleobase stacking such as intercalators or nucleobase-binding groups.

In the above-indicated probes, one or more of the Q-groups may designate a label. Examples of suitable labels are given below. Moieties wherein Q denotes a label may preferably be located in one or both of the terminating moieties of the probe. Moieties wherein Q denotes a label may, however, also be located internally.

The peptide nucleic acid probes may comprise moieties, wherein all X groups are O (polyamides) or wherein all X groups are S (polythioamides). It is to be understood that the probes may also comprise mixed moieties (comprising both amide and thioamide moieties).

In another aspect, the present invention relates to peptide nucleic acid probes of formula (II), (III) and (IV) as well as mixtures of such probes according to claim 9.

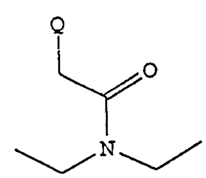
In a preferred embodiment, the peptide nucleic acid probes or mixtures thereof according to the invention are of formulas (I)-(IV) as defined in claim 10 with Z being NH, NCH_3 or O, each R^2, R^3 and R^4 independently being H or the side chain of a naturally occurring amino acid, the side chain of a non-naturally occurring amino acid, or C_{1-4} alkyl, and each Q being a naturally occurring nucleobase or a non-naturally occurring nucleobase with the provisos defined in

claims 4 to 8.

Peptide nucleic acid probes or mixtures of such probes according to the invention are preferably those of formula (I)-(IV) as defined in claim 11 with Z being NH or O, and R^2 being H or the side chain of Ala, Asp, Cys, Glu, His, HomoCys, Lys, Orn, Ser or Thr, and Q being a nucleobase selected from thymine, adenine, cytosine, guanine, uracil, iso-C, and 2,6-diaminopurine with the provisos defined in claims 4 to 8.

Peptide nucleic acid probes or mixtures thereof, which are of major interest for detecting mycobacteria of the MTC group or one or more mycobacteria other than mycobacteria of the MTC group, are probes of formula (V) according to claim 12, wherein R^4 is H or the side chain of Ala, Asp, Cys, Glu, His, HomoCys, Lys, Orn, Ser or Thr, Q is as defined in claim 11 and with the provisos indicated in claims 4 to 8.

The peptide nucleic acid probe comprises polymerised moieties as defined above and in the claims. From the formula, it is to be understood that the probe may comprise polymerised moieties which structure may be mutually different or identical. It may be advantageous that at least one moiety of the probe, preferably one (or both) of the moieties terminating the probe, are of a different structure. Such terminating moieties may suitably be a moiety of formula (VI)



(VI)

where Q is as defined above. Such moiety may suitably be connected to a peptide nucleic acid moiety though an amide bond.

The peptide nucleic acid probe according to the invention comprises from 6 to 30 polymerised moieties of formulas (I) to (V), and, in addition, optionally one or two terminating moieties of formula (VI) as defined above. The preferred length of the probe will depend on the sample material and whether labelled probes are used. It is contemplated that especially interesting probes comprise from 10 to 30 polymerised moieties of formulas (I) to (V), and, in addition, optionally one or two terminating moieties of formula (VI) as defined above. Probes of the invention may suitably comprise from 12 to 25 polymerised moieties of formulas (I) to (V), more suitably from 14 to 22 polymerised moieties of formulas (I) to (V), most suitably from 15 to 20 polymerised moieties of formulas (I) to (V), and, in addition, optionally one or two terminating moieties of formula (VI).

In many cases, it may be advantageous to use a mixture of probes, e.g. probes of different length and/or probes directed to different domains in the rRNA. Probes labelled with different labels may also be applied, thus allowing differentiation between the probes. Thereby, mycobacteria of the MTC group and one or more mycobacteria other than those of the MTC group may be detected simultaneously.

As mentioned above, the polymerised moieties of the probes may be mutually different or identical. In some embodiments, the polymerised moieties of formulas (V) constitute at least 75% by weight (calculated by excluding labels and linkers), preferably at least 80% by weight and most preferably at least 90% by weight of the probe.

The ends on the moieties terminating the probe may be substituted by suitable substituents which in the following will be named "linkers". A terminating end may suitably be substituted by from 1 to 5 linkers, more suitably from 1 to 3 linkers. Such linkers may suitably be selected among C_{1-15} alkyl, C_{1-15} alkenyl and C_{1-15} alkynyl groups as defined above. The linkers may be substituted or unsubstituted, branched or non-branched, or be interrupted by heteroatoms, or be substituted or contain functional groups as described above. This may depend on the chemical nature of the terminating moiety (i.e. whether the moiety is terminated by a carbon, oxygen or nitrogen atom). A terminating end or a linker on a terminating end may further be substituted by one or more labels, which labels may be incorporated end to end, i.e. so as to form a non-branched labelled end, or may be incorporated so as to form a branched labelled end ("zipper"). The linkers may be attached directly to a terminating end, may be attached to a label or between labels on a terminating end, or be attached to a terminating end before a label is attached to a terminating end. It should be understood that two terminating ends may carry different or identical substituents, linkers and/or labels. It should further be understood that the term "a label" is intended to comprise one or more labels as the term "linkers" is to comprise one or more linkers. For certain applications, it may be advantageous that one or more linkers are incorporated between the peptide nucleic acid moieties. Such applications may in particular be those based on triplex formation.

Examples of suitable linkers are $-NH(CH_2CH_2O)_nCH_2C(O)-$, $-NH(CHOH)_nC(O)-$, $-(O)C(CH_2OCH_2)_nC(O)-$ and $-NH(CH_2)_nC(O)-$, $NH_2(CH_2CH_2O)_nCH_2C(O)-$, $NH_2(CHOH)_nC(O)-$, $HO(O)C(CH_2OCH_2)_nC(O)-$, $NH_2(CH_2)_nC(O)-$, $-NH(CH_2CH_2O)_nCH_2C(O)OH$, $-NH(CHOH)_nC(O)OH$, $-(O)C(CH_2OCH_2)_nC(O)OH$ and $-NH(CH_2)_nC(O)OH$, wherein n is 0 or an integer from 1 to 8, preferably from 1 to 3. Examples of very interesting linkers are $-NHCH_2C(O)-$, $-NHCH_2CH_2C(O)-$, $-NH(CH_2CH_2O)_2CH_2C(O)-$, $HO(O)CCH_2CH_2C(O)(NH-(CH_2CH_2O)_2CH_2C(O))_2-$.

In the present context, the term "label" refers to a substituent which is useful for detection or visualisation. Suitable labels comprise fluorophores, biotin, dinitro benzoic acid, digoxigenin, radioisotope labels, peptide or enzyme labels, chemiluminiscence labels, hapten, antigen or antibody labels.

The expression "peptide label" is intended to mean a label comprising from 1 to 20 naturally occurring or non-naturally occurring amino acids, preferably from 1 to 10 naturally occurring or non-naturally occurring amino acids, more preferably from 1 to 8 naturally occurring or non-naturally occurring amino acids, most preferably from 1 to 4 naturally occurring or non-naturally occurring amino acids, linked together end to end in a non-branched or branched ("zipper") fashion. Such peptide label may be composed of amino acids which are mutually identical or different. In a preferred embodiment, such a non-branched or branched end comprises one or more, preferably from 1 to 8 labels, more preferably from 1 to 4, further labels other than a peptide label. Such further labels may suitably terminate a non-branched end or a branched end. One or more linkers may suitably be attached to the terminating end before a peptide label and/or a further label is attached. Such linker units may also be attached between a peptide label and a further label. Furthermore, such peptide labels may be incorporated between the peptide nucleic acid moieties.

The probe as such may also comprise one or more labels such as from 1 to 8, preferably from 1 to 4, labels and/or one or more linker units, which may be attached internally, i.e. to the backbone of the probe. The linker units and labels may mutually be attached as described above.

Examples of particular interesting labels are biotin, fluorescent labels, such as fluorescein labels, e.g. 5-(and 6)-carboxyfluorescein, 5- or 6-carboxyfluorescein, 6-(fluorescein)-5-(and 6)-carboxamido hexanoic acid and fluorescein isothiocyanate, peptide labels consisting of from 1 to 20 naturally occurring amino acids or non-naturally occurring amino acids, peroxidases such as horse radish peroxidase (HRP) and soya bean peroxidase, dinitro benzoic acid, rhodamine, tetramethylrhodamine, cyanine dyes such as Cy2, Cy3 and Cy5, coumarin, R-phycoerythrin (RPE), allophycoerythrin, Texas Red and Princeton Red as well as conjugates of R-phycoerythrin and, e.g. Cy5 or Texas Red.

Examples of preferred labels are biotin, fluorescent labels, peptide labels and dinitro benzoic acid. Peptide labels may preferably be composed of from 1 to 10, more preferably of from 1 to 8, most preferably of from 1 to 4, naturally occurring or non-naturally occurring amino acids. It may be particularly advantageous to incorporate one or more other labels as well as a peptide label such as from 1 to 8 or from 1 to 4 other labels, preferably one or two other labels.

Suitable peptide labels may preferably be composed of cysteine, glycine, lysine or ornithine.

In a further embodiment, the Q substituent as defined above may be labelled. Suitable labels are as defined above. Between Q and such a label, a linker as defined above may be incorporated. It is preferred that such labelled ligands Q are selected from thymine and uridine labelled in the 5-position and 7-deazaguanine and 7-deazaadenine labelled in the 7-position.

A mixture of peptide nucleic acids is also part of the present invention. Such mixture may comprise more than one probe capable of hybridising to 23S rRNA, more than one probe capable of hybridising to 16S rRNA, or more than one probe capable of hybridising to 5S rRNA as well as more than one probe capable of hybridising to 23S rRNA, and/or more than one probe capable of hybridising to 16S rRNA, and/or more than one probe capable of hybridising to 5S rRNA. The mixture may also comprise peptide nucleic acids for detecting more than one mycobacteria in the same assay.

The probes may be synthesised according to the procedures described in "PNA Information Package" obtained from Millipore Corporation (Bedford, MA, USA), or may be synthesised on an Expedite Nucleic Acid Synthesis System (PerSeptive BioSystems, USA).

If using the Fmoc strategy for elongation of the probe with linkers or amino acids, it is possible to retain side chain amino groups protected with acid sensitive protection groups such as the Boc or Mtt group. This method allows introduction of a linker containing several Boc protected amino groups which can all be cleaved and labelled in the same synthesis cycle.

One way of labelling a probe is to use a fluorescent label, such as 5-(and 6)-carboxyfluorescein, 5- or 6-carboxyfluorescein, or 6-(fluorescein)-5-(and 6)-carboxamido hexanoic acid. The acid group is activated with HATU (O-(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate) or HBTU (2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate) and reacted with the N-terminal amino group of the peptide nucleic acid. The same technique can be applied to other labelling groups containing an acid function. Alternatively, the succinimidyl ester of the above-mentioned labels or fluorescein isothiocyanate may be used directly.

After synthesis, probes can be cleaved from the resin using standard procedures as described by Millipore Corporation or PerSeptive Biosystems. The probes are subsequently purified and analysed using reversed-phase HPLC techniques at 50°C and were characterised by matrix-assisted laser desorption/ionisation time of flight mass spectrometry (MALDI-TOFMS), plasma

desorption mass spectrometry (PDMS), electron spray mass spectrometry (ESMS), or fast atom bombardment (FAB-MS).

Generally, probes such as probes comprising polymerised moieties of formula (IV) and (V) may also be prepared as described in, e.g., Angewandte Chemie, International Edition in English 35, 1939-1942 (1996) and Bioorganic & Medical Chemistry Letters, Vol 4, No 8, 1077-1080 (1994). Chemical properties of some probes are described in, e.g., Nature, 365, 566-568 (1993).

Detection of the label depend on the type of label and on the format of the procedure. In cases where the sample is deposited onto slides, the hybridisation results may be visualised using well known immunohistochemical staining methods for detection of the labelling on the probe. When fluorescent labelled probes are used, the hybrids may be detected using an antibody against the fluorescent label which antibody may be conjugated with an enzyme. The fluorescent label may alternatively be detected directly using a fluorescence microscope, or the results may be automatically analysed on a fluorescent-based image analysis system.

When biotin labelled probes are used, the hybrids may be detected using streptavidin or an antibody against the biotin label which antibody or streptavidin may be conjugated with an enzyme. If necessary, an enhancement of the signal can be generated using commercially available amplification systems such as the catalysed signal amplification system for biotinylated probes (e.g. DAKO K 1500).

The probes according to the invention are used in the detection of mycobacteria, in particular mycobacteria of the MTC group or one or more mycobacteria other than mycobacteria of the MTC group, in samples, in particular sputum samples, which samples may contain these bacteria.

In the assay method according to the invention, a sample to be analysed for the presence of mycobacteria is brought into contact with one or more probes according to the invention under conditions by which hybridisation between the probe and any sample rRNA or rDNA originating from mycobacteria can occur, and the resulting hybridisation is observed or measured.

If necessary, a mixture of random probes (probes with random, not selected sequences optionally of different length) may suitably be added in excess in admixture with the probe to reduce non-specific binding.

In one embodiment of the assay method, one or more probes according to the invention are brought into contact with any target rRNA or rDNA inside the cells (in situ) under suitable stringency conditions. Prior to this step, smears of the bacterial cells are prepared using conventional procedures. Following hybridisation, the complexes formed are detected. An example of this assay format is fluorescence *in situ* hybridisation (FISH), wherein the probes according to the invention are labelled with fluorescein or another fluorophore. It might be advantageous to use more than one probe. If e.g. three such probes are included in the assay each in a concentration of one third of the concentration of a single probe, possible cross reactivity of the individual probes is less likely to invalidate the results.

The test sample may prior to hybridisation be subjected to conditions which release RNA from the organisms present in the sample. Such RNA may in solution be brought into contact with one or more probes as defined herein, which optionally are labelled, or the RNA of the test sample may be immobilised onto a solid phase prior to hybridisation with one or more detection probes. The RNA may be immobilised by dotting the RNA onto membranes or by using an immobilised capture probe.

The sample comprising the target nucleic acid can even be added to an assay system comprising detection probes as well as immobilised capture probe. The immobilisation of the capture probe may be effected by using a streptavidin coated solid phase and a biotinylated capture probe. The capture probe may further be immobilised onto a solid support by coupling reaction between a carboxylic acid on the linker and an amino derivatised support. Alternatively, the coupling onto the solid support may be accomplished by photochemical activation of photoreactive groups which have been attached absorptively to the solid support prior to photochemical activation. Such photoreactive groups are described in EP 408 078 A.

In practice, a solid phase based assay system is very attractive as the analysis can be carried out using a solid phase precoated with a capture probe. A solid phase based assay system is also feasible for automatisation of the analysis.

The capture probe may be one of the other probes not used in the hybridisation reaction and detection step for target nucleic acid, thus ensuring dual species specificity. The dual specificity will allow shorter probes be used, e.g. 10 mer probes.

Furthermore, the capture of purine rich sequences may be improved by utilising bis-peptide nucleic acids as capture probes. Such bis-peptide nucleic acids are described in WO 96/02558. The bis-peptide nucleic acids comprise a first peptide nucleic acid strand capable of hybridising in parallel fashion to the target nucleic acid, and a second peptide nucleic acid

strand capable of hybridising in antiparallel fashion to the purine rich sequence of the nucleic acid to be captured. The two peptide nucleic acid strands are connected by a linker and are in this way capable of forming a triplex structure with said purine rich sequence nucleic acid.

The number of polymerised moieties of each linker-separated peptide nucleic acid may be as previously defined for non-bis-peptide nucleic acids. However, due to the high stability of the triplexes formed, bis-peptide nucleic acids with short first and second strands can be used which will make the design of a pyrimidine rich probe easier.

The solid support capture system may take a wide variety of forms well known in the art, such as e.g. a plate, a microtiter plate having one or more wells, a microscope slide, a filter, a membrane, a tube, a dip stick, a strip, beads such as paramagnetic beads, beads made of polystyrene, polypropylene, polyethylene, dextran, nylon, amyloses, natural and modified celluloses, polyacrylamides and agaroses. When a filter, a membrane, a strip or beads is (are) used as the solid support, it (they) may, if conveniently be incorporated into a single-use device.

It has been observed that peptide nucleic acids may bind to a variety of solid phases. A blocking reaction may thus be required to reduce non-specific binding of the peptide nucleic acids to the solid phase. The blocking reaction may be carried out with commonly used blocking reagents, such as BSA (bovine serum albumin), casein, Triton X-100® or Tween 20®. The preferred blocking reagents are Triton X-100® and Tween 20®.

The captured probe:nucleic acid complexes may be detected or identified by a wide variety of methods for that purpose. The probe to be brought in contact with the target nucleic acid may be labelled, and the label may be detected using well known detection systems. In another embodiment, the captures probe:nucleic acid complexes may be detected using a detection system based on an antibody reacting specifically with complexes formed between peptide nucleic acid and nucleic acid (such as described in WO 95/17430), in which detection system the primary antibody may comprise a label, or which detection system comprises a labelled secondary antibody, which specifically binds to the primary antibody.

The present probes further provide a method of diagnosing infection by mycobacteria and a method for determining the stage of the infection and the appropriate treatment by which methods one or more optionally labelled probes according to the invention are brought into contact with a patient sample and the type of treatment and/or the effect of a treatment is (are) evaluated.

DESCRIPTION OF SPECIFIC EMBODIMENTS

Examples of suitable Qs of adjacent moieties are given below. Peptide nucleic acid probes comprising such Qs will be suitable for detecting mycobacteria, in particular mycobacteria of the MTC group or mycobacteria other than mycobacteria of the MTC group. The probes are written from left to right corresponding to from the C-terminal end towards the N-terminal end. Suitable Q subsequences for detecting 23S and 16S mycobacterial rRNA of the MTC group are disclosed in Danish patent applications DK 1096/96 and DK 1156/96. Other suitable Q subsequences for detecting 23S and 16S rRNA as well as 5S rRNA of the MTC group are given below. Suitable Q subsequences for detecting 23S and 16S rRNA of mycobacteria other than mycobacteria of the MTC group are given below.

MTC group (23S)

Suitable Q subsequences directed to 23S rRNA of mycobacteria of the MTC group are given below.

AGA TGC GGG TAG CAC (selected from position 149-158 in Figure 1A)
TGT TTT CTC CTC CTA (selected from position 220-221 in Figure 1A)
ACT GCC TCT CAG CCG (selected from position 328-361 in Figure 1B)
TGA TAC TAG GCA GGT (selected from position 453-455 in Figure 1B)
CGG ATT CAC AGC GGA (selected from position 490-501 in Figure 1C
TCA CCA CCC TCC TCC (selected from position 637-660 in Figure 1C)
TTA ACC TTG CGA CAT (selected from position 706-712 in Figure 1D)
ACT ATT CAC ACG CGC (selected from positions 762-789 in Figure 1D)
CTC CGC GGT GAA CCA (selected from position 989 in Figure 1D)
GCT TTA CAC CAC GGC (selected from position 1068-1072 in Figure 1E)
ACG CTT GGG GGC CCT (selected from position 1148 in Figure 1E)
CCA CAC CCA CCA CAA (selected from position 1311-1329 in Figure 1F)
CCG GTG GCT TCG CTG (selected from position 1361-1364 in Figure 1F)
ACT TGC CTT GTC GCT (selected from position 1418 in Figure 1G)
GAT TCG TCA CGG GCG (selected from position 1563-1570 in Figure 1G)
AAC TCC ACA CCC CCG (selected from position 1627-1638 in Figure 1G)
ACC CCT TCG CTT GAC (selected from position 1675-1677 in Figure 1H)
CTT GCC CCA GTG TAA (selected from position 1718 in Figure 1H)
CTC TCC CTA CCG GCT (selected from position 1734-1740 in Figure 1H)
GAT ATT CCG GTC CCC (selected from position 1967-1976 in Figure 1I)
ATC CCG CCC CAA CTG (selected from position 2403-2420 in Figure 1I)
CTG TCC CTA AAC CCG (selected from position 2457-2488 in Figure 1J)
TTC GAG GTT AGA TGC (selected from position 2457-2488 in Figure 1J)

GGT GCA CCA GAG GTT (selected from position 2952-2956 in Figure 1J)
CTG GCG GGA CAA CTG (selected from position 2966-2969 in Figure 1K)
TTA TCC TGA CCG AAC (selected from position 3000-3003 in Figure 1K)
GAC CTA TTG AAC CCG (selected from position 3097-3106 in Figure 1K)

MTC group (16S)
GAA GAG ACC TTT CCG (selected from position 76-79 in Figure 2A)
CAC TCG AGT ATC TCC (selected from position 98-101 in Figure 2A)
ATC ACC CAC GTG TTA (selected from position 136-136 in Figure 2A)
GCA TCC CGT GGT CCT (selected from position 194-201 in Figure 2B)
GCA TCC CGT GGT CCT (selected from position 194-201 in Figure 2B)
TAA AGC GCT TTC CAC (selected from position 222-229 in Figure 2B)
GCT CAT CCC ACA CCG (selected from position 242 in Figure 2B)
CCG AGA GAA CCC GGA (selected from position 474 in Figure 2C)
AGT CCC CAC CAT TAC (selected from position 1136-1145 in Figure 2C)
AAC CTC GCG GCA TCG (selected from position 1271-1272 in Figure 2C)
GGC TTT TAA GGA TTC (selected from position 1287-1292 in Figure 2D)
GAC CCC GAT CCG AAC (selected from position 1313 in Figure 2D)
CCG ACT TCA CGG GGT (selected from position 1334 in Figure 2D)

MTC group (5S)
Suitable Q subsequences directed to 5S rRNA of mycobacteria of the MTC group are given below.

CGG AGG GGC AGT ATC (selected from position 86-90 in Figure 3)

Mycobacteria other than those of the MTC group (23S)
Suitable Q subsequences directed to 23S rRNA of mycobacteria other than those of the MTC group are given below.
GAT CAA TGC TCG GTT (selected from position 99-101 in Figure 4A)
TTC CCC GCG TTA CCT (selected from position 183 in Figure 4A)
TTA GCC TGT TCC GGT (selected from position 261-271 in Figure 4B)
GCA TGC GGT TTA GCC (selected from position 281-284 in Figure 4B)
TAC CCG GTT GTC CAT (selected from position 290-293 in Figure 4B)
GTA GAG CTG AGA CAT (selected from position 327-335 and 343-357 in Figure 4C)
GCC GTC CCA GGC CAC (selected from position 400-405 in Figure 4C)
CTC GGG TGT TGA TAT (selected from position 453-462 in Figure 4D)
ACT ATT TCA CTC CCT (selected from position 587-599 in Figure 4D)

ACG CCA TCA CCC CAC (selected from position 637-660 in Figure 4E)
CGA CGT GTC CCT GAC (selected from position 704-712 in Figure 4E)
ACT ACA CCC CAA AGG (selected from position 763-789 in Figure 4F)
CAC GCT TTT ACA CCA (selected from position 1060-1074 in Figure 4F)
GCG ACT ACA CAT CCT (selected from position 1177-1185 in Figure 4F)
CGG CGC ATA ATC ACT (selected from position 1259-1265 in Figure 4G)
CCA CAT CCA CCG TAA (selected from position 1311-1327 in Figure 4G)
CGC TGA ATG GGG GAC (selected from position 1345-1348 in Figure 4G)
GGA GCT TCG CTG AAT (selected from position 1361-1364 in Figure 4H)
CGG TCA CCC GGA GCT (selected from position 1361-1364 in Figure 4H)
GGA CGC CCA TAC ACG (selected from position 1556-1570 in Figure 4H)
GAA GGG GAA TGG TCG (selected from position 1608-1613 in Figure 4I)
AAT CGC CAC GCC CCC (selected from position 1626-1638 in Figure 4I)
CAG CGA AGG TCC CAC (selected from position 1651-1659 in Figure 4I)
GTC ACC CCA TTG CTT (selected from position 1675-1677 in Figure 4I)
ATC GCT CTC TAC GGG (selected from position 1734-1741 in Figure 4I)
GTG TAT GTG CTC GCT (selected from position 1847-1853 in Figure 4J)
ACG GTA TTC CGG GCC (selected from position 1967-1976 in Figure 4J)
GGC CGA ATC CCG CTC (selected from position 2006-2010 in Figure 4J)
AAA CAG TCG CTA CCC (selected from position 2025-2027 in Figure 4J)
CCT TAC GGG TTA ACG (selected from position 2131-2132 in Figure 4K)
GAG ACA GTT GGG AAG (selected from position 2252-2255 in Figure 4K)
TGG CGT CTG TGC TTC (selected from position 2396-2405 in Figure 4 L)
CGA CTC CAC ACA AAC (selected from position 2416-2420 in Figure 4L)
GAT AAG GGT TCG ACG (selected from position 2474-2478 in Figure 4L)
ATC CGT TGA GTG ACA (selected from position 2687 in Figure 4M)
CAG CCC GTT ATC CCC (selected from position 2719 in Figure 4M)
AAC CTT TGG GAC CTG (selected from position 2809 in Figure 4M)
TAA AAG GGT GAG AAA (selected from position 3062-3068 in Figure 4N)
GTC TGG CCT ATC AAT (selected from position 3097-3106 in Figure 4N)

Mycobacteria other than those of the MTC group (16S)

Suitable Q subsequences directed to 16S rRNA of mycobacteria other than those of the MTC group are given below.

AGA TTG CCC ACG TGT (selected from position 135-136 in Figure 5A)
AAT CCG AGA AAA CCC (selected from position 472-475 in Figure 5A)
GCA TTA CCC GCT GGC (selected from position 1136-1144 in Figure 5A)
TTA AAA GGA TTC GCT (selected from position 1287-1292 in Figure 5B)

AGA CCC CAA TCC GAA (selected from position 1313 in Figure 5B)

GAC TCC GAC TTC ATG (selected from position 1334 in Figure 5B)

The invention is further illustrated by the non-limiting examples given below.

EXAMPLES

EXAMPLE 1

In situ hybridisation to fixed bacterial cells

To test the ability of the peptide nucleic acid probes to detect MTC and not mycobacteria other than MTC, in particular not mycobacteria of the M. avium Complex, or Neisseria gonorrhoeae, fluorescence *in situ* hybridisation (FISH) was performed on fixed bacterial cells using fluorescein labelled probes as shown below. It was shown that these probes did not hybridise to M. avium, M. intracellulare, or N. gonorrhoeae.

Preparation of bacterial slides

M. bovis BCG (Statens Seruminstitut, Denmark, Catalogue number 2645), M. avium (Statens Seruminstitut, Denmark, Laboratory number 3716 (E37978)), and M. intracellulare (Statens Seruminstitut, Laboratory number 3717 (E39562)) were grown in Dubos medium (Statens Seruminstitut, Denmark) or on Löwenstein-Jensen medium (Statens Seruminstitut, Denmark) at 37 °C. N. gonorrhoeae was grown on chocolate agar at 37 °C with additional 5% CO_2.

Bacterial smears were prepared on test slides according to standard procedures. The smears were air-dried followed by flame fixation.

FISH on bacterial slides

The following procedure was performed.

1. The slide is immersed in 80% ethanol for 15 minutes, subsequently rinsed with water and air-dried.
2. The bacterial slide is covered with a hybridisation solution containing the probe in question.
3. The slide is incubated in a humid incubation chamber at 45°C or 55°C for 90 minutes.
4. The slide is washed 25 minutes in TBS-buffer, pH 10 at 45°C or 55°C, followed by 30 seconds in water.
5. The slide is dried and mounted (DAKO Fluorescence Mounting Medium or equivalent).

The following hybridisation solutions was used:

Hybridisation solution	10 mM NaCl
	10% Dextran sulphate
	30% formamide
	0.1% Triton X-100®
	50 mM Tris-HCl, pH 7.6
	50 mM EDTA
	0.1% sodium pyrophosphate
	0.2% polyvinylpyrrolidone
	0.2% Ficol

TBS buffer	10 mM sodium phosphate, pH 10
	145 mM NaCl

All solutions are made RNase free following standard procedures.

The following peptide nucleic acid probes were used

Lys(Flu)-Lys(Flu)-TTC GAG GTT AGA TGC-NH_2	OK 306
Lys(Flu)-Lys(Flu)-ACT CCA CAC CCC CGA-NH_2	OK 309
Lys(Flu)-Lys(Flu)-TCA CCA CCC TCC TCC-NH_2	OK 446
Lys(Flu)-Lys(Flu)-AAC TCC ACA CCC CCG-NH_2	OK 449
Lys(Flu)-Lys(Flu)-TCA CCA CCC TCC TCC-NH_2	OK 447
Lys(Flu)-Lys(Flu)-TTC GAG GTT AGA TGC-NH_2	OK 306
Lys(Flu)-Lys(Flu)-CAC AAG ACA TGC ATC-NH_2	OK 310

wherein Flu denotes a fluorescein isothiocyanate label or a 5-(and 6)-carboxyfluorescein label, and Lys(Flu)-Lys(Flu) denotes a peptide label ("zipper") with two Flu labels attached. All the probes are directed to 23S rRNA of the mycobacteria of the MTC group, except OK 310 which is directed to 16S rRNA. The results are shown in Table 1.

TABLE 1

	OK 306 (250nM) 45°C	OK 309 (250nM) 45°C	OK 446 (500nM) 55°C	OK 449 (500nM) 55°C
M. bovis BCG	positive	positive	positive	positive
M. avium	negative	negative	negative	negative
M. intracellulare	negative	negative	not determined	not determined
N. gonorrhoeae	negative	negative	not determined	not determined

	OK 447 (1μM) 55°C	OK 310 (250nM) 45°C	OK 306 (500nM) OK 310 (500nM) 55°C
M. bovis BCG	positive	positive	positive
M. avium	negative	negative	negative
M. intracellulare	not determined	negative	negative
N. gonorrhoeae	not determined	negative	not determined

EXAMPLE 2

Test in dot blots

To further test the ability of the peptide nucleic acid probes to detect MTC and not MAC or E. coli, dot blot tests were carried out.

M. bovis BCG (Statens Seruminstitut Catalogue number 2645) and M. intracellulare (Statens Seruminstitut, Denmark Laboratory number 3713 (E39562)) were grown in Dubos medium (Statens Seruminstitut, Denmark) or on Löwenstein-Jensen medium (Statens Seruminstitut, Denmark) at 37 °C.

RNA was isolated from the bacterial cells by use of TRI-reagent (Sigma) following manufacture's directions. E. coli rRNA was purchased from Boehringer Mannheim, Germany.

The following peptide nucleic acid probes were used.

Lys(Flu)-Lys(Flu)-CTG TCC CTA AAC CCG-NH_2 OK 305
Lys(Flu)-Lys(Flu)-GTC CCT AAA CCC GAT-NH_2 OK 307
Lys(Flu)-Lys(Flu)-ACT CCA CAC CCC CGA-NH_2 OK 309
Lys(Flu)-Lys(Flu)-Gly-GCA TCC CGT GGT CCT-NH_2 OK 223
Lys(Flu)-Lys(Flu)-CAC AGG ACA TGC ATC-NH_2 OK 310

wherein Flu denotes a fluorescein isothiocyanate label or a 5-(and 6)-carboxyfluorescein label, and Lys(Flu)-Lys(Flu) denotes a peptide label ("zipper") consisting of 2 amino acids, respectively, with two Flu labels attached. OK 305, OK 307 and OK 309 are directed to 23S rRNA of the mycobacteria of the MTC group. OK 223 and OK 310 are directed to 16S rRNA of the mycobacteria of the MTC group.

Preparation of dot blots

The following buffers were used:

20 × SSPE buffer	3 M NaCl
	0.2 M PO_4^{3-}
	0.02 M EDTA
	pH 7.4
TST buffer	0.05 M Tris/HCl
	0.5 M NaCl
	0.5% Tween 20®
	pH 9.0

200 ng M. bovis RNA, M. intracellulare RNA and E. coli rRNA were dotted onto membranes (Schleich & Schuel, NY 13 N), and the membranes were dried and fixed under UV light for 2 minutes. Each of the probes (70 nM probe in hybridisation solution (hybridisation solution without Triton X-100® and with the exception that formamide was substituted with 50% glycerol)) were added to the membrane. Hybridisation was continued for 1.5 hours at 55 °C. The membranes were rinsed 2 times for 15 minutes in 2 × SSPE buffer containing 0.1% SDS at ambient temperature, and subsequently 2 times for 15 minutes in 0.1 × SSPE buffer containing 0.1% SDS at 55 °C or at 65 °C (see Table 2). The membrane was blocked with 0.5% casein dissolved in 0.5M NaCl, 0.05M Tris/HCl pH 9.0. Thereafter, the membranes were incubated for 1 hour with rabbit-anti FITC antibody labelled with AP (DAKO K0046 vial A) diluted 1:2000 in 0.5% casein dissolved in 0.5M NaCl, 0.05M Tris/HCl pH 9.0. After incubation, the membranes were washed 3 times 5 minutes with TST at ambient temperature. Bound probes were visualised following standard procedures using BCIP/NBT, and the visualisation was stopped by incubation for 10 minutes with 10 mM EDTA. The blot was dried at 50 °C.

The results are given in Table 2 below.

TABLE 2

	E. coli rRNA		M. bovis BCG RNA		M. intracellulare RNA	
Probe	55 °C	65 °C	55 °C	65 °C	55 °C	65 °C
OK 305	negative	negative	positive	positive	negative	weak
OK 307	negative	negative	positive	positive	negative	weak
OK 309	negative	negative	positive	positive	negative	weak
OK 223	negative	negative	positive	positive	nd	nd
OK 310	negative	negative	negative	positive	negative	negative

nd: Not determined

EXAMPLE 3

Test of probes on clinical smears of sputum

The ability of the peptide nucleic acid to bind to mycobacteria of the MTC group was tested on clinical smears of sputum prepared by the Division of Microbiology at Ramathibodi Hospital, Bangkok, Thailand. Smears from the same patient were initially evaluated positive by Ziehl-Neelsen staining. Ziehl-Neelsen staining only shows the presence of acid fast bacilli, not whether these are mycobacteria of the MTC group.

The following probes were used.

Lys(Flu)-Lys(Flu)-TCA CCA CCC TCC TCC-NH_2 OK 446
Lys(Flu)-Lys(Flu)-AAC TCC ACA CCC CCG-NH_2 OK 449
Lys(Flu)-Lys(Flu)-TTC GAG GTT AGA TGC-NH_2 OK 306
Lys(Flu)-Lys(Flu)-CAC AAG ACA TGC ATC-NH_2 OK 310

wherein Flu denotes a fluorescein isothiocyanate label or a 5-(and 6)-carboxyfluorescein label, and Lys(Flu)-Lys(Flu) denotes a peptide label ("zipper") consisting of 2 amino acids with two Flu labels attached. OK 446, OK 449 and OK 306 are directed to 23S rRNA of MTC, whereas OK 310 is directed towards 16S rRNA of MTC. Furthermore, a random peptide nucleic acid probe (a 15-mer wherein each position may be A, T, C or G obtained from Millipore Corporation (Bedford, MA, USA)) was used in order to increase the signal-to-noise ratio.

The clinical smears were prepared according to the procedure described in Example 1. The probe concentrations were varied as indicated below in the Table 3. The results are shown in

Table 3.

TABLE 3

Sample number	OK 446 (1μM) Random (50μM)	OK 449 (1μM) Random (50μM)	Ziehl-Neelsen staining
285	Positive	Positive	4+
335	Positive	Eq.	2+
345	Positive	Positive	3+
224	Positive	Positive	3+
297	Negative	Eq.	2+
179	Negative	Negative	4+
247	Negative	Negative	2+
255	Positive	Positive	2+
202	Eq.	Positive	2+

Sample number	OK 306 (500nM) OK 310 (500nM)	Ziehl-Neelsen staining
213	Positive	4+
292	Positive	4+
159	Positive	3+
287	Positive	3+

Smears stained by Ziehl-Neelsen staining were examined with a 100x objective and scored according to the following method: -: 0 baccilli, +/-: 1-200 per 300 fields, 2+: 1-9 per 10 fields, 3+: 1-9 per field, 4+: >9 per field.

Smears using FISH were examined with a 100x objective and scored according to the following method: Positive: Several mycobacteria were identified in the smear. Negative: No fluorescent mycobacteria were identified in the smear. Eq: Few (1-3) fluorescent mycobacteria were identified in the smear.

It appears from the table that some of the Ziehl-Neelsen positive smears are MTC-negative. The results thus indicate that the peptide nucleic acid probes can separate MTC-positive and MTC-negative samples.

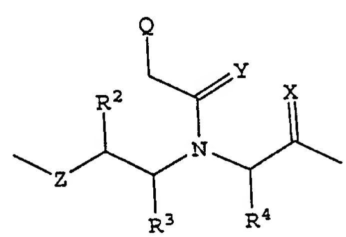
CLAIMS

1. Peptide nucleic acid probe for detecting the presence of mycobacteria in a sample, said probe being capable of hybridising to target sequences in rRNA or DNA from the area coding for said rRNA of the mycobacteria forming detectable target-probe hybrids, and a mixture of such probes.

2. Peptide nucleic acid probe according to claim 1, said probe being capable of hybridising to 23S, 16S or 5S rRNA or DNA from the area coding for said rRNA of the mycobacteria forming detectable target-probe hybrids, and a mixture of such probes.

3. Peptide nucleic acid probe according to claim 1 or 2 for detecting the presence of mycobacteria of the Mycobacterium tuberculosis Complex (MTC), in particular M. tuberculosis, or one or more mycobacteria other than mycobacteria of the Mycobacterium tuberculosis Complex (MTC), in particular mycobacteria of the Mycobacterium avium Complex, in a sample, which probe comprises from 6 to 30 polymerised peptide nucleic acid moieties, said probe being capable of hybridising to target sequences in 23S, 16S or 5S rRNA or DNA from the area coding for said rRNA of the mycobacteria to be detected, and a mixture of such probes.

4. Peptide nucleic acid probe according to any one of claims 1 to 3 for detecting mycobacteria of the Mycobacterium tuberculosis Complex (MTC) in a sample, which probe comprises from 10 to 30 polymerised moieties of formula (I)



(I)

wherein each X and Y independently designate O or S, each Z independently designates O, S, NR^1, or $C(R^1)_2$, wherein each R^1 independently designate H, C_{1-6} alkyl, C_{1-6} alkenyl, C_{1-6} alkynyl, each R^2, R^3 and R^4 designate independently H, the side chain of a naturally occurring amino acid, the side chain of a non-naturally occurring amino acid, C_{1-4} alkyl, C_{1-4} alkenyl or C_{1-4} alkynyl, or a functional group, each Q independently designates a naturally occurring nucleobase, a non-naturally occurring nucleobase, an intercalator, a nucleobase-binding

group, a label or H,

with the proviso that the Qs of adjacent moieties are selected so as to form a sequence of which a subsequence includes at least one nucleobase complementary to a nucleobase of M. tuberculosis 23S rRNA that differs from the corresponding nucleobase of M. avium located within the following domains

Positions 149-158 in Figure 1A,
Positions 220-221 in Figure 1A,
Positions 328-361 in Figure 1B,
Positions 453-455 in Figure 1B,
Positions 490-501 in Figure 1C,
Positions 637-660 in Figure 1C,
Positions 706-712 in Figure 1D,
Positions 762-789 in Figure 1D,
Position 989 in Figure 1D,
Positions 1068-1072 in Figure 1E,
Position 1148 in Figure 1E,
Positions 1311-1329 in Figure 1F,
Positions 1361-1364 in Figure 1F,
Position 1418 in Figure 1G,
Positions 1563-1570 in Figure 1G,
Positions 1627-1638 in Figure 1G,
Positions 1675-1677 in Figure 1H,
Position 1718 in Figure 1H,
Positions 1734-1740 in Figure 1H,
Positions 1967-1976 in Figure 1I,
Positions 2403-2420 in Figure 1I,
Positions 2457-2488 in Figure 1J,
Positions 2952-2956 in Figure 1J,
Positions 2966-2969 in Figure 1K,
Positions 3000-3003 in Figure 1K or
Positions 3097-3106 in Figure 1K,

and further with the proviso that the probe comprising such subsequence is able to form hybrids with target sequences in 23S rRNA of said mycobacteria,
and a mixture of such probes.

5. Peptide nucleic acid probe according to any one of claims 1 to 3 for detecting mycobacteria of the Mycobacterium tuberculosis Complex (MTC) in a sample, which probe comprises from 10 to 30 polymerised moieties of formula (I) as defined in claim 4,

with the proviso that the Qs of adjacent moieties are selected so as to form a sequence of which a subsequence includes at least one nucleobase complementary to a nucleobase of M. tuberculosis 16S rRNA that differs from the corresponding nucleobase of M. avium located within the following domains

Positions 76-79 in Figure 2A,
Positions 98-101 in Figure 2A,
Positions 135-136 in Figure 2 A,
Positions 194-201 in Figure 2B,
Positions 222-229 in Figure 2B,
Position 242 in Figure 2B,
Position 474 in Figure 2C,
Positions 1136-1145 in Figure 2C,
Positions 1271-1272 in Figure 2C,
Positions 1287-1292 in Figure 2D,
Position 1313 in Figure 2D, or
Position 1334 in Figure 2D,

and further with the proviso that the probe comprising such subsequence is able to form hybrids with target sequences in 16S rRNA of said mycobacteria, and a mixture of such probes.

6. Peptide nucleic acid probe according to any one of claims 1 to 3 for detecting mycobacteria of the Mycobacterium tuberculosis Complex (MTC) in a sample, which probe comprises from 10 to 30 polymerised moieties of formula (I) as defined in claim 4,

with the proviso that the Qs of adjacent moieties are selected so as to form a sequence of which a subsequence includes at least one nucleobase complementary to a nucleobase of M. tuberculosis 5S rRNA that differs from the corresponding nucleobase of M. avium located within the following domain

Positions 86-90 in Figure 3

and further with the proviso that the probe comprising such subsequence is able to form

hybrids with target sequences in 5S rRNA of said mycobacteria, and a mixture of such probes.

7. Peptide nucleic acid probe according to any one of claims 1 to 3 for detecting one or more mycobacteria other than mycobacteria of the Mycobacterium tuberculosis Complex in a sample, which probe comprises from 10 to 30 polymerised moieties of formula (I) as defined in claim 4,

with the proviso that the Qs of adjacent moieties are selected so as to form a sequence of which a subsequence includes at least one nucleobase complementary to a nucleobase of M. avium 23S rRNA that differs from the corresponding nucleobase of M. tuberculosis located within the following domains

Positions 99-101 in Figure 4A,
Position 183 in Figure 4A,
Positions 261-271 in Figure 4B,
Positions 281-284 in Figure 4B,
Positions 290-293 in Figure 4B,
Positions 327-335 in Figure 4C,
Positions 343-357 in Figure 4C,
Positions 400-405 in Figure 4C,
Positions 453-462 in Figure 4D,
Positions 587-599 in Figure 4D,
Positions 637-660 in Figure 4E,
Positions 704-712 in Figure 4E,
Positions 763-789 in Figure 4F,
Positions 1060-1074 in Figure 4F,
Positions 1177-1185 in Figure 4F,
Positions 1259-1265 in Figure 4G,
Positions 1311-1327 in Figure 4G,
Positions 1345-1348 in Figure 4G,
Positions 1361-1364 in Figure 4H,
Positions 1556-1570 in Figure 4H,
Positions 1608-1613 in Figure 4I,
Positions 1626-1638 in Figure 4I,
Positions 1651-1659 in Figure 4I,
Positions 1675-1677 in Figure 4I,
Positions 1734-1741 in Figure 4I,

Positions 1847-1853 in Figure 4J,
Positions 1967-1976 in Figure 4J,
Positions 2006-2010 in Figure 4J,
Positions 2025-2027 in Figure 4J,
Positions 2131-2132 in Figure 4K,
Positions 2252-2255 in Figure 4K,
Positions 2396-2405 in Figure 4L,
Positions 2416-2420 in Figure 4L,
Positions 2474-2478 in Figure 4L,
Position 2687 in Figure 4M,
Position 2719 in Figure 4M,
Position 2809 in Figure 4M,
Positions 3062-2068 in Figure 4N, or
Positions 3097-3106 in Figure 4N,

and further with the proviso that the probe comprising such subsequence is able to form hybrids with target sequences in 23S rRNA of said mycobacteria,
and a mixture of such probes.

8. Peptide nucleic acid probe according to any one of claims 1 to 3 for detecting one or more mycobacteria other than mycobacteria of the Mycobacterium tuberculosis Complex in a sample, which probe comprises from 10 to 30 polymerised moieties of formula (I) as defined in claim 4,

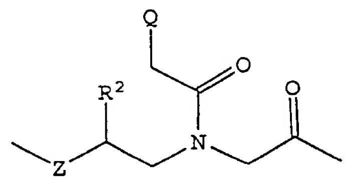
with the proviso that the Qs of adjacent moieties are selected so as to form a sequence of which a subsequence includes at least one nucleobase complementary to a nucleobase of M. avium 16S rRNA that differs from the corresponding nucleobase of M. tuberculosis located within the following domains

Positions 135-136 in Figure 5A,
Positions 472-475 in Figure 5A,
Positions 1136-1144 in Figure 5A,
Positions 1287-1292 in Figure 5B,
Position 1313 in Figure 5B, or
Position 1334 in Figure 5B,

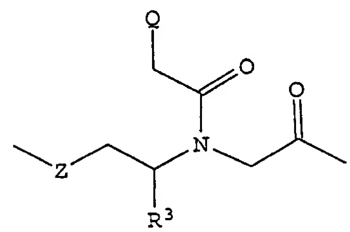
and further with the proviso that the probe comprising such subsequence is able to form hybrids with target sequences in 16S rRNA of said mycobacteria,

and a mixture of such probes.

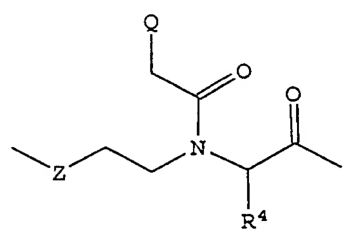
9. Peptide nucleic acid probe according to any one of claims 1 to 8 of formula (II), (III), or (IV)



(II)



(III)



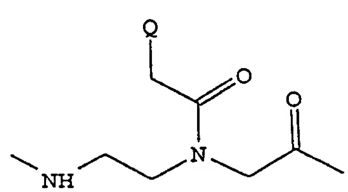
(IV)

wherein Z, R^2, R^3, and R^4, and Q is as defined in claim 4, and a mixture of such probes.

10. Peptide nucleic acid probe according to any one of claims 1 to 9, wherein Z is NH, NCH_3 or O, each R^2, R^3 and R^4 independently designate H or the side chain of a naturally occurring amino acid, the side chain of a non-naturally occurring amino acid, or C_{1-4} alkyl, and each Q is a naturally occurring nucleobase or a non-naturally occurring nucleobase with the provisos defined in claims 4 to 8, and a mixture of such probes.

11. Peptide nucleic acid probe according to any one of claims 1 to 10, wherein Z is NH or O, and R^2 is H or the side chain of Ala, Asp, Cys, Glu, His, HomoCys, Lys, Orn, Ser or Thr, and Q is a nucleobase selected from thymine, adenine, cytosine, guanine, uracil, iso-C and 2,6-diaminopurine with the provisos defined in claims 4 to 8, and a mixture of such probes.

12. Peptide nucleic acid probe according to any one of claims 1 to 11 of formula (V)



(V)

wherein R^4 is H or the side chain of Ala, Asp, Cys, Glu, His, HomoCys, Lys, Orn, Ser or Thr, and Q is as defined in claim 11 with the provisos defined in claims 4 to 8, and a mixture of such probes.

13. Peptide nucleic acid probe according to any one of claims 1 to 12 further comprising one or more labels and a mixture of such probes according to any one of claims 1 to 12 further comprising one or more labels which may be mutually identical or different, which probes optionally may comprise one or more linkers which may be mutually identical or different with the provisos defined in claims 4 to 8.

14. Peptide nucleic acid probes according to any one of claims 1 to 6 and 8 to 13, wherein the Qs of adjacent moieties are selected so as to form the following subsequences

TTC GAG GTT AGA TGC,	(OK 306)
ACT CCA CAC CCC CGA,	(OK 309)
TCA CCA CCC TCC TCC,	(OK 446)
AAC TCC ACA CCC CCG,	(OK 449)
TCA CCA CCC TCC TCC,	(OK 447)
TTC GAG GTT AGA TGC,	(OK 306)
CAC AAG ACA TGC ATC,	(OK 310)
CTG TCC CTA AAC CCG,	(OK 305)
GTC CCT AAA CCC GAT, or	(OK 307)
GCA TCC CGT GGT CCT,	(OK 223)

and mixtures of such probes.

15. A mixture of peptide nucleic acid probes according to claims 1 to 14.

16. Use of a peptide nucleic acid probe or a mixture thereof according to any one of claims 1 to 15 for the detection of the presence of mycobacteria in a sample.

17. Use of a peptide nucleic acid probe or a mixture thereof according to any one of claims 1 to 6 and 9 to 15 for the detection of the presence of mycobacteria of the Mycobacterium tuberculosis Complex (MTC), in particular M. tuberculosis.

18. Use of a peptide nucleic acid probe or a mixture thereof according to any one of claims 1 to 3, 7 to 13 and 15 for the detection of the presence of one or more mycobacteria other than mycobacteria of the Mycobacterium tuberculosis Complex, in particular mycobacteria of the

Mycobacterium avium Complex.

19. Method for detecting mycobacteria in a sample comprising

(1) contacting any rRNA or rDNA optionally present in said sample with one or more peptide nucleic acid probes according to anyone of claims 1 to 15 under conditions, whereby hybrids between said probe(s) and said rRNA are formed, and

(2) observing or measuring said hybridisation, and relating said observation or measurement to the presence of mycobacteria in said sample.

20. Method according to claim 19 for detecting mycobacteria of the Mycobacterium tuberculosis Complex (MTC), in particular M. tuberculosis.

21. Method according to claim 19 for detecting one or more mycobacteria other than mycobacteria of the Mycobacterium tuberculosis Complex.

22. Method according to any one of claims 19 to 21, wherein the contact between any rRNA optionally present in the sample and one or more peptide nucleic acid probes according to anyone of claims 1 to 14 or mixtures thereof according to claim 15 takes place inside the cells (in situ).

23. Method according to any one of claims 19 to 21,
c h a r a c t e r i s e d in that the hybridisation is performed in solution and the hybrids are captured on a solid phase before measuring the extent of hybridisation.

24. Method according to claim 23,
c h a r a c t e r i s e d in that a peptide nucleic acid probe according to any one of claims 1 to 15 are used for capturing the hybrids.

25. Method according to any one of claims 19 to 21,
c h a r a c t e r i s e d in that RNA of the test sample is immobilised onto a solid phase prior to performing step (1).

26. A method according to any one of claims 19 to 25,
c h a r a c t e r i s e d in that a signal amplifying system is used for measuring the resulting hybridisation.

27. Kit for detecting mycobacteria, in particular mycobacteria of the Mycobacterium tuberculosis Complex (MTC), in particular M. tuberculosis, and/or one or more mycobacteria other than mycobacteria of the Mycobacterium tuberculosis Complex, in particular mycobacteria of the Mycobacterium avium Complex,
c h a r a c t e r i s e d in that said kit comprises at least one peptide nucleic acid probe according to any one of claims 1 to 14, and optionally a detection system with at least one detecting reagent.

28. Kit according to claim 27,
c h a r a c t e r i s e d in that it further comprises a solid phase capture system.

ABSTRACT

NOVEL PROBES FOR THE DETECTION OF MYCOBACTERIA

Novel hybridisation assay probes and mixtures of such probes for detecting the presence of rRNA originating from mycobacterial species. The probes may suitable be directed to 23S, 16S or 5S rRNA of said mycobacteria. Such probes are capable of detecting the organisms in test samples, e.g. expectorates, sputum, aspirates, urine, blood and tissue sections, food, soil and water.

```
                    130       140       150       160
1093 GGGGAAACCCAGCACGAGTGATGTCGTGCTACCCGCATCT M.tuberculosis
422  GGGGGAACCCAGCACGAGTGATGTCGTGTTACCCGTATCT M.avium
422  GGGGGAACCCAGCACGAGTGATGTCGTGTTACCCGTATCT M.paratuberc.
507  GGGGGAACCCGGCACGAGTGATGTCGTGTCACCCAACGCT M.phlei
432  GGGGAAACCCAACACGAGTCACGTCGTGTTACCCGTATCT M.leprae
207  GGGGAAACCCAGCACGAGTAATGTCGTGTTACCCGTATCT M.gastri
150  GGGGAAACCCAGCACGAGTGATGTCGTGTTACCCGCATCT M.kansasii

2588 GGGGAAACCCGGCACGAGTGATGTCGTGTCACCAGGCGCT M.smegmatis
```

• • •

```
                    210       220       230       240
1172 CATCTCAGTACCCGTAGGAGCAGAAAACAATTGTGATTCC M.tuberculosis
501  CATCTCAGTACCCGTAGGAGAAGAAAACAATTGTGATTCC M.avium
501  CATCTCAGTACCCGTAGGAGAAGAAAACAATTGTGATTCC M.paratuberc.
586  CATCTCAGTACCCGTAGAAGAAGAAAACAATTGTGATTCC M.phlei
511  CATCTCAGTACCCGTAGGAGAAGAAAACAATTGTGATTCC M.leprae
286  CATCTCAGTACCCGTAGGAGAAGAAAACAAAAGTGATTCC M.gastri
229  CATCTCAGTACCCGTAGGAGAAGAAAACAAAAGTGATTCC M.kansasii

2667 CATCTCAGT-CCCGTAGGAAGAGAAAACAAATGTGATTCC M.smegmatis
```

Figure 1A

	330 340 350 360	
1289	TGTGGAG-GATATGTCTCAGCGCTACCCGGCTGAGA-GG	M.tuberculosis
617	TGTGGATTGATATGTCTCAGCTCTACCTGGCTGAGG-GG	M.avium
617	TGTGGATTGATATGTCTCAGCTCTACCTGGCTGAGG-GG	M.paratuberc.
703	TGTGGGCCCTGTGTGTC-CATCGTCCGCCGGCGATGGCAG	M.phlei
629	TGTGGATTGGTATGTCTCAACTCTACCTGGTTGAGG-GG	M.leprae
404	TGTGGATCGATACGTCTCAGCTCTACCCGGCTGAGG-GG	M.gastri
347	TGTGGATCGATACGTCTCAGCTCTACCCGGCTGAGG-GG	M.kansasii
2785	TGTGGACCTATCTTTC-CGCCTCTACCTGGCTG-GAGGG	M.smegmatis

	370 380 390 400	
1327	CAGTCAGAAAGTGTCGTGGTTAGCGGAAGTGGCCTGGGAT	M.tuberculosis
656	TAGTCAGAAAGTGTCGTGGTTAGCGGAAGTGGCCTGGGAC	M.avium
656	TAGTCAGAAAGTGTCGTGGTTAGCGGAAGTGGCCTGGGAC	M.paratuberc.
742	TAGTGATAAAGCAGTGTGGTTAGGTGAAGTGGCCTGGGAT	M.phlei
668	TAGTCAGAAAGTGCCGTGGTTAGCGGAAATGGCCTGGGAT	M.leprae
443	CAGTCAGAAAGTGTCGTGGTTAACGGAAGTGGCCTGGGAT	M.gastri
386	CAGTCAGAAAGTGTCGTGGTTAACGGAAGTGGCCTGGGAT	M.kansasii
2823	CAGTGAGAAAATGTTGTGGTTAGCGGAAATGGCTTGGGAT	M.smegmatis

• • •

	450 460 470 480	
1406	CGGCACCTGCCTAGTATCAATTCCCGAGTAGCAGCGGGCC	M.tuberculosis
735	CGGCACCTGCCTTATATCAACACCCGAGTAGCAGCGGGCC	M.avium
735	CGGCACCTGCCTTATATCAACACCCGAGTAGCAGCGGGCC	M.paratuberc.
820	TGCTGCC-GCTGTCACAGG--TCCCGAGTAGCAGCGGGCC	M.phlei
747	TGGCACCTGCCTTGTATCAATTCCCGAGTAGCAGCGGGCC	M.leprae
522	CGGCACCTGCCTTGTATCAATTCCCGAGTAGCAGCGGGCC	M.gastri
465	CGGCACCTGCCTTGTATCAATTCCCGAGTAGCAGCGGGCC	M.kansasii
2902	CGACGTCTGTCTTGATGGTGTTCCCGAGTAGCAGCGGGCC	M.smegmatis

Figure 1B

```
                   490         500         510         520
1446 CGTGGAATCCGCTGTGAATCCGCCGGGACCACCCGGTAAG M.tuberculosis
775  CGTGGAATCTGCTGTGAATCTGCCGGGACCACCCGGTAAG M.avium
775  CGTGGAATCTGCTGTGAATCTGCCGGGACCACCCGGTAAG M.paratuberc.
857  CGTGGAATCTGCTGTGAATCTGCCGGGACCACCCGGTAAG M.phlei
787  CGTGGAATCTGCTGTGAATCTGCCGGGACCACCCGGTAAG M.leprae
562  CGTGGAATCTGCTGTGAATCTGCCGGGACCACCCGGTAAG M.gastri
505  CGTGGAATCTGCTGTGAATCTGCCGGGACCACCCGGTAAG M.kansasii

2942 CGTGGAATCTGCTGTGAATCTGCCGGGACCACCCGGTAAG M.smegmatis
```

• • •

```
                   610         620         630         640
1566 GTACCTGAAACCGTGTGCCTACAATCCGTCAGAGCCTCCT M.tuberculosis
894  GTACCTGAAACCGTGTGCCTACAATCCGTCAGAGCCTCCT M.avium
894  GTACCTGAAACCGTGTGCCTACAATCCGTCAGAGCCTCCT M.paratuberc.
976  GTACCTGAAACCGTGTGCCTACAATCCGTCAAAGCCCTCT M.phlei
907  GTACCTGAAACCGTGTGCCTACAATCCGTCAGAGCCTCTT M.leprae
682  GTACCTGAAACCGTGTGCCTACAATCCGTCAGAGCCCCTT M.gastri
625  GTACCTGAAACCGTGTGCCTACAATCCGTCAGAGCCCTTT M.kansasii

3062 GTACCTGAAACCGTGCGCTTACAATCCGTCAGAGCCCTCG M.smegmatis

                   650         660         670         680
1606 TTTCCTCTCCGGAGGAGGGTGGTGATGGCGTGCCTTTTGA M.tuberculosis
934  C---------------GTGGGGTGATGGCGTGCCTTTTGA M.avium
934  C---------------GTGGGGTGATGGCGTGCCTTTTGA M.paratuberc.
1016 CTT----------GTAGTGGGGTGATGGCGTGCCTTTTGA M.phlei
947  T---------------GTGGGGTGATGGCGTGCTTTTTGA M.leprae
722  T---------------GTGGGGTGATGGCGTGCCTTTTGA M.gastri
665  C---------------GTGGGGTGATGGCGTGCCTTTTGA M.kansasii

3102 ACGTGT----------GTGGGGTGATGGCGTGCCTTTTGA M.smegmatis
```

Figure 1C

```
                 690       700       710       720
1646 AGAATGAGCCTGCGAGTCAGGGACATGTCGCAAGGTTAAC M.tuberculosis
959  AGAATGAGCCTGCGAGTCAGGGACACGTCGCGAGGTTAAC M.avium
959  AGAATGAGCCTGCGAGTCAGGGACACGTCGCGAGGTTAAC M.paratuberc.
1046 AGAATGAGCCTGCGAGTCAGGGACATGTCGCGAGGTTAAC M.phlei
972  AGAATGAGCCTGCGAGTCAGGGACATGTCGCGAGGTTAAC M.leprae
747  AGAATGAGCCTGCGAGTCAGGGACATGTCGCGAGGTTAAC M.gastri
690  AGAATGAGCCTGCGAGTCAGGGACATGTCGCGAGGTTAAC M.kansasii

3132 AGAATGAGCCTGCGAGTCAGGGACATGTCGCGAGGTTAAC M.smegmatis
```

• • •

```
                 770       780       790       800
1726 CGACCCACACGCGCATACGCGCGTGTGAATAGTGGCGTGT M.tuberculosis
1039 CGCATCCCCTTTGCG-----------GTGTAGTGGCGTGT M.avium
1039 CGCATCCCCTTTGCG-----------GTGTAGTGGCGTGT M.paratuberc.
1126 CGTATCCAACCTGTT----GGCGTTGGTGTAGTGGTGTGT M.phlei
1052 CGTATCACGTGTGAGCGT--------GTGTAGTGGCGTGT M.leprae
827  CGTATCACGCGTAAGCGT--------GTGTAGTGGCGTGT M.gastri
770  CGTATCGCGCGCGAGCGT--------GTGTAGTGGCGTGT M.kansasii

3212 CGTATCCACACAAGAGTGTGTG----GTGTAGTGGTGTGT M.smegmatis
```

• • •

```
                 970       980       990       1000
1926 ATTTAGGTGCAGCGTTGCGTGGTTCACCGCGGAGGTAGAG M.tuberculosis
1228 ATTTAGGTGCAGCGTTGCGTGGTTCACCACGGAGGTAGAG M.avium
1228 ATTTAGGTGCAGCGTTGCGTGGTTCACCACGGAGGTAGAG M.paratuberc.
1322 ATTTAGGTGCAGCGTCGCATGTTTCTTATCGGAGGTAGAG M.phlei
1244 ATTTAGGTGCAGCGTTGCGTGGTTCACCACGGAGGTAGAG M.leprae
1019 ATTTAGGTGCAGCGTTGCGTGTTTCACCACGGAGGTAGAG M.gastri
962  ATTTAGGTGCAGCGTTGCGTGTTTCACCACGGAGGTAGAG M.kansasii

3408 ATTTAGGTGCAGCGTCGCATGTTTCTTGCCGGAGGTAGAG M.smegmatis
```

Figure 1D

```
            1050      1060      1070      1080
2005 CAGCCAAACTCCGAATGCCG-TGGTG-TA-AAGCGTGGCA M.tuberculosis
1307 CAGCCAAACTCCGAATGCCG-TGGTG-TAAAAGCGTGGCA M.avium
1307 CAGCCAAACTCCGAATGCCG-TGGTG-TAAAAGCGTGGCA M.paratuberc.
1401 CAGCCAAACTCCGAATGCCGATAAG--TGAAAGTGTGGCA M.phlei
1323 CAGCCAAACTCCGAATGCCG-TGGTT-TAAAAGCGTGGCA M.leprae
1098 CAGCCAAACTCCGAATGCCG-TGGTG-TATA-GCGTGGCA M.gastri
1041 CAGCCAAACTCCGAATGCCG-TGGTG-TATA-GCGTGGCA M.kansasii

3486 CAGCCAAACTCCGAATGCCGGTAAGGCCAAGAGTGCGGAA M.smegmatis
```

• • •

```
            1130      1140      1150      1160
2082 ACAGCCCAGATCGCCGGCTAAGGCCCCCAAGCGTGTGCTA M.tuberculosis
1385 ACAGCCCAGATCGCCGGCTAAGGCCCCTAAGCGTGTGCTA M.avium
1385 ACAGCCCAGATCGCCGGCTAAGGCCCCTAAGCGTGTGCTA M.paratuberc.
1479 ACAGCCCAGATCGCCGGCTAAGGCCCCTAAGCGTGTGCTA M.phlei
1401 ACAGCCCAGATCGCCGGCTAAGGCCCCTAAGCGTGTGCTA M.leprae
1175 ACAGCCCAGATCGCCGGCTAAGGCCCCAAAGCGTGTGCTA M.gastri
1118 ACAGCCCAGATCGCCGGCTAAGGCCCCAAAGCGTGTGCTA M.kansasii

3566 ACAGCCCAGATCGCCGGTTAAGGCCCCTAAGCGTTTGTTA M.smegmatis
```

Figure 1E

```
                    1290      1300      1310      1320
2241 CTCAAGCACACCGCCGAAGCCGCGGCACATCCACCTTGT- M.tuberculosis
1544 CTCAAGCACACCGCCGAAGCCGCGGCACATTCATCTT-TA M.avium
1544 CTCAAGCACACCGCCGAAGCCGCGGCACATTCATCTT-TA M.paratuberc.
1638 CTCAAGCACACCGCCGAAGCCGCGGCA--ATCAGCGTTTG M.phlei
1560 CTCAAGCACACCGCCGAAGCCGCGGCACATTCACCTTCTA M.leprae
1334 CTCAAGCACACCGCCGAAGCCGCGACA-----ACCGC--A M.gastri
1277 CTCAAGCACACCGCCGAAGCCGCGACA-----ACCGC--A M.kansasii

3726 TTCAAGCACACCGCCGAAGCCGCGGAA--GCCAACGTTTG M.smegmatis

                    1330      1340      1350      1360
2280 -GGTGGGTGTGGGTAGGGGAGCGTCCCTCATTCAGCGAAG M.tuberculosis
1583 CGGTGGATGTGGGTAGGGGAGCGTCCCCCATTCAGCGAAG M.avium
1583 CGGTGGATGTGGGTAGGGGAGCGTCCCCCATTCAGCGAAG M.paratuberc.
1676 TGGCTGGTGTGGGTAGGGGAGCGTCCTGCATCCGGTGAAG M.phlei
1600 GGGTGGATGTGGGTAGGGGAGCGTTCCTCATTCAGCGAAG M.leprae
1367 AGGT-----TGGGTAGGGGAGCGTCCCTCATTCAGCGAAG M.gastri
1310 AGGT-----TGGGTAGGGGAGCGTCCCTCATTCAGCGAAG M.kansasii

3764 TT-------TGGGTAGGGGAGCGTCCTG-ATCCGGTGAAG M.smegmatis

                    1370      1380      1390      1400
2319 CCACCGGGTGACCGGTGGTGGAGGGTGGGGGAGTGAGAAT M.tuberculosis
1623 CT-CCGGGTGACCGGTGGTGGAGGGTGGGGGAGTGAGAAT M.avium
1623 CT-CCGGGTGATCGGTGGTGGAGGGTGGGGGAGTGAGAAT M.paratuberc.
1716 CCGCCGAGTGATCGGTGGTGGAGGGTGTGGGAGTGAGAAT M.phlei
1640 CCTCCGGGTAACCGGTGGTGGAGGGTGGGGAAGTGAGAAT M.leprae
1402 CCGCCGGGTGACCGGTGGTGGAGGATGGGGGAGTGAGAAT M.gastri
1345 CTGCCGGGTGACCGGTGGTGGAGGATGGGGGAGTGAGAAT M.kansasii

3796 CCGCCGAGTATCGAGTGGTGGAGGGTGTGGGAGTGAGAAT M.smegmatis
```

Figure 1F

```
                 1410      1420      1430      1440
2359 GCAGGCATGAGTAGCGACAAGGCAAGTGAGAACCTTGCCC M.tuberculosis
1662 GCAGGCATGAGTAGCGATAAGGCAAGTGAGAACCTTGCCC M.avium
1662 GCAGGCATGAGTAGCGATAAGGCAAGTGAGAACCTTGCCC M.paratuberc.
1756 GCAGGCATGAGTAGCGATAAGGCAAGTGAGAACCTTTCCC M.phlei
1680 GCAGGCATGAGTAGCGATAAGGCAAGTGAGAACCTTGCCC M.leprae
1442 GCAGGCATGAGTAGCGATAAGGCAAGTGAGAACCTTGCCC M.gastri
1385 GCAGGCATGAGTAGCGATAAGGCAAGTGAGAACCTTGCCC M.kansasii

3836 GCAGGCATGAGTAGCGATTAGGCAAGTGAGAACCTTCCCC M.smegmatis

• • •

                 1570      1580      1590      1600
2519 CGCCCGTGACGAATCA-GCGGTACTAACCACCCAAAACCG M.tuberculosis
1821 CGTCCCTGATGAATCA-GCGGTACTAACCACCCAAAACCG M.avium
1821 CGTCCCTGATGAATCA-GCGGTACTAACCACCCAAAACCG M.paratuberc.
1915 CGTCCCTGATGAATCTCATTCTGCTAACCACCCAAAACCT M.phlei
1840 CGCCCGTGATGAATCA-GCGGTACTCACCACCCAAAACCG M.leprae
1602 CGCCCGTGATGAATCA-GCGGTACTAACCACCCAAAACCG M.gastri
1545 CGCCCGTGATGAATCA-GCGGTACTAACCACCCAAAACCG M.kansasii

3996 CGTCCATGATGAATCA-GCGGTACTAACCATCCAAAACCA M.smegmatis

                 1610      1620      1630      1640
2558 GAT-CGATCAC-TCCCCTTCGGGGG-TGTGGAGTTC-TGG M.tuberculosis
1860 GAT-CGACCAT-TCCCCTTCGGGGC-GTGGCGATT-CGG M.avium
1860 GAT-CGACCAT-TCCCCTTCGGGGC-GTGGCGATT-CGG M.paratuberc.
1955 GGC-CGATC--ATCC--TTCGGGG--GTGACGGTTG-GG M.phlei
1879 GAT-CGACCATATCCCCTTCGGGGCTATGGAGGTT-CGG M.leprae
1641 GAT-CGATCAC-TCCCCTTCGGGGA-GTGGAGGTC-TGG M.gastri
1584 GAT-CGATCAC-TCCCCTTCGGGGC-GTGGAGGTC-TGG M.kansasii

4035 ACCGTGACCGCACCT--TTCGGGG--TGTGGCGTTGGTGG M.smegmatis
```

Figure 1G

```
                1650      1660      1670      1680
2594 GGCTGCGTGGGAACTTCGCTGGTAGTAGTCAAGCGAAGGG M.tuberculosis
1896 GGCTGCGTGGGACCTTCGCTGGTAGTAGTCAAGCAATGGG M.avium
1896 GGCTGCGTGGGACCTTCGCTGGTAGTAGTCAAGCAATGGG M.paratuberc.
1986 GGCTGCGTGGGACCCG-GTGGGTAGTAGTCAAGCGATGGG M.phlei
1917 GGCTGCGTGGGAACTTCGTTGGTAGTAGTCAAGCGATGGG M.leprae
1677 GGCTGCGTGGAGCCTTCGCTGGTAGTAGTCAAGCGATGGG M.gastri
1620 GGCTGCGTGGAGCCTTCGCTGGTAGTAGTCAAGCGATGGG M.kansasii

4071 GGCTGCATGGGACCTTCGTTGGTAGTAGTCAAGCGATGGG M.smegmatis

                1690      1700      1710      1720
2634 -GTGACGCAGGAAGGTAGCCGTACCAGTCAGTGGTAACA- M.tuberculosis
1936 -GTGACGCAGGAAGGCAGCCGTACCAGTCAGTGGTAATA- M.avium
1936 -GTGACGCAGGAAGGCAGCCGTACCAGTCAGTGGTAATA- M.paratuberc.
2025 -GTGACGCAGGAAGGTAGCCGTACCAGTCAGTGGTAATA- M.phlei
1957 -GTGACGCAGGAAGGTAGCCGTACCAGTCAGTGGTAATA- M.leprae
1717 -GTGACGCAGGAAGGCAGCCGTACCAGTCAGTGGTAATA- M.gastri
1660 -GTGACGCAGGAAGGCAGCCGTACCAGTCAGTGGTAATA- M.kansasii

4111 -GTGACGCAGGAAGGTAGCCGTACCGGTCAGTGGTAATA- M.smegmatis

                1730      1740      1750      1760
2672 -CTGGGGCAAGCCGGTAGGGAGAGCGATAGGCAAATCCGT M.tuberculosis
1974 -CTGGGGCAAGCCCGTAG--AGAGCGATAGGCAAATCCGT M.avium
1974 -CTGGGGCAAGCCCGTAG--AGAGCGATAGGCAAATCCGT M.paratuberc.
2063 -CCGGGGTAAACCTGTAGGGCGAGTGATAGGCAAATCCGT M.phlei
1995 -CTGGAGCAAGCCCGTAGGGAGAGCGATAGGCAAATCCGT M.leprae
1755 -CTGGGGCAAGCCAGTAGGGAGAGCGATAGGCAAATCCGT M.gastri
1698 -CTGGGGCAAGCCAGTAGGGAGAGCGATAGGCAAATCCGT M.kansasii

4149 -CCGGCGTAAGCCTGTAGGGAGTCAGATAGGTAAATCCGT M.smegmatis
```

Figure 1H

```
                 1970      1980      1990      2000
2908 AGGGGGACCGGAATATCGTGAACACCCTTGCGGTGGGAGC M.tuberculosis
2208 AGGGGGCCCGGAATACCGTGAACACCCTTGCGGTGGGAGC M.avium
2208 AGGGGGCCCGGAATACCGTGAACACCCTTGCGGTGGGAGC M.paratuberc.
2298 AGGGGGACCCACGTACCGTGAGGGCTCTTGCGGCGGCAGC M.phlei
2231 AGGGGGGCCGGAATATCGTGAACACCCTTGCGGTGGGAGC M.leprae
1910                                          M.gastri
1934 AGGGGGACCGGAATACCGTGAACACCCTTGCGGTGGGAGC M.kansasii

4385 AGGGGGACCCACATGGCGTGTAAGCCTTTACGGCCCAAGC M.smegmatis
```

• • •

```
                 2410      2420      2430      2440
3345 ACCTCGACGCCAGTTGGGGCGGAGTCGTTGTTGAAATACC M.tuberculosis
2645 GCACAGACGCCAGTTTGTGTGGAGTCGTTGTTGAAATACC M.avium
2645 GCACAGACGCCAGTTTGTGTGGAGTCGTTGTTGAAATACC M.paratuberc.
2737 GCTCGGACGCCAGTTCGGGTGGAGTCGTTGTTGAAATACC M.phlei
2668 ACTTCGACGCTAGTTGGGGTGGAGTCGTTGTTGAAATACC M.leprae
1910                                          M.gastri
2372 ACCTCAACGCCAGTTGGGGTGGAGTCGTTGTTGAAATACC M.kansasii

4822 GCTCACACGCCAGTGTGGGTGGAGTCGTTGTTGAAATACC M.smegmatis
```

Figure 1I

```
               2450      2460      2470      2480
3385 ACTCTGATCGTATTGGGCATCTAACCTCGAACCCTGAATC M.tuberculosis
2685 ACTCTGATCGTATTGGACACCTAACGTCGAACCCT-TATC M.avium
2685 ACTCTGATCGTATTGGACACCTAACGTCGAACCCT-TATC M.paratuberc.
2777 ACTCTGATCGTATTGGGCCTCTAACCTCGGACCGTGGATC M.phlei
2708 ACTCTGATTGTATTGAACATCTAACCTCGAACCGTATATC M.leprae
1910                                          M.gastri
2412 ACTCTGATCGTATTGGACACCTAACGTCGAACCCTGAATC M.kansasii

4862 ACTCTGATCGTATTGGGCCTCTAACCTCGGACCGTATATC M.smegmatis

               2490      2500      2510      2520
3425 GGGTTTAGGGACAGTGCCTGGCGGGTAGTTTAACTGGGGC M.tuberculosis
2724 GGGTTCACGGACAGTGCCTGGCGGGTAGTTTAACTGGGGC M.avium
2724 GGGTTCACGGACAGTGCCTGGCGGGTAGTTTAACTGGGGC M.paratuberc.
2817 CGGTTCAGGGACAGTGCCTGGTGGGTAGTTTAACTGGGGC M.phlei
2748 CGGTTTAGGGACAGTGCCTGGCGGGTAGTTTAACTGGGGC M.leprae
1910                                          M.gastri
2452 GGGTTCACGGACAGTGCCTGGCGGGTAGTTTAACTGGGGC M.kansasii

4902 CGGTTCAGGGACAGTGCCTGGTGGGTAGTTTAACTGGGGC M.smegmatis
```

• • •

```
               2930      2940      2950      2960
3864 AGTACGAGAGGACCGGGACGGACGAACCTCTGGTGCACCA M.tuberculosis
3163 AGTACGAGAGGACCGGGACGGACGAACCTCTGGTATACCA M.avium
3163 AGTACGAGAGGACCGGGACGGACGAACCTCTGGTATACCA M.paratuberc.
3256 AGTACGAGAGGACCGGGACGGACGAACCTCTGGTATACCA M.phlei
3187 AGTACGAGAGGACCGGGACGGACGAACCTCTGGTATACCA M.leprae
1910                                          M.gastri
2891 AGTACGAGAGGACCGGGACGGACGAACCTCTAGTGCACCA M.kansasii

5342 AGTACGAGAGGACCGGGACGGACGAACCTCTGGTATACCA M.smegmatis
```

Figure 1J

```
                 2970      2980      2990      3000
3904 GTTGTCCCGCCAGGGGCACCGCTGGATAGCCACGTTCGGT M.tuberculosis
3203 GTTGTCCCACCAGGGGCACGGCTGGATAGCCACGTTCGGA M.avium
3203 GTTGTCCCACCAGGGGCACGGCTGGATAGCCACGTTCGGA M.paratuberc.
3296 GTTGTCCCACCAGGGGCACCGCTGGATAGCCACGTTCGGA M.phlei
3227 GTTGTCTCACCAGGGGCACCGCTGGATAGCCACGTTCGGA M.leprae
1910                                          M.gastri
2931 GTTGTCCCACCAGGGGCACCGCTGGATAGCTACGTTCGGA M.kansasii

5382 GTTGTCCCACCAGGGGCACGGCTGGATAGCCACGTTCGGA M.smegmatis

                 3010      3020      3030      3040
3944 CAGGATAACCGCTGAAAGCATCTAAGCGGGAAACCTTCTC M.tuberculosis
3243 CAGGATAACCGCTGAAAGCATCTAAGCGGGAAACCTTCTC M.avium
3243 CAGGATAACCGCTGAAAGCATCTAAGCGGGAAACCTTCTC M.paratuberc.
3336 CAGGATAACCGCTGAAAGCATCTAAGCGGGAAACCTCTTC M.phlei
3267 CAAGATAACCGCTGAAAGCATCTAAGCGGGAAACCTTCTC M.leprae
1910                                          M.gastri
2971 CAGGATAACCGCTGAAAGCATCTAAGCGGGAAACCTTCTC M.kansasii

5422 CAGGATAACCGCTGAAAGCATCTAAGCGGGAAACCTCTTC M.smegmatis
```

• • •

```
                 3090      3100      3110      3120
4023 CCCGC-AGAACACGGGTTCAATAGGTCAGACCTGGAAGCT M.tuberculosis
3322 CCCGC-AGACCACGGGATTGATAGGCCAGACCTGGAAGCT M.avium
3322 CCCGC-AGATCACGGGATTGATAGGCCAGACCTGGAAGCT M.paratuberc.
3415 CCCGC-AGACCACGGGATCGATAGACCAGACCTGCACGCA M.phlei
3309                                          M.leprae
1910                                          M.gastri
3050 CCCGC-AGAACACGGGTTCGATAGGCCAGACCTGGAAGCT M.kansasii

5501 CCCGC-AGACCACGGGATTGATAGACCAGACCTGGAAGCC M.smegmatis
```

Figure 1K

```
                          50        60        70        80
2     GCGGCGTGCTTAACACATGCAAGTCGAACGGAAAGGTCTC M.tuberculosis
141   GCGGCGTGCTTAACACATGCAAGTCGAACGGAAAGGTCTC M.bovis
39    GCGGCGTACTTAACACATGCAAGTCGAACGGAAAGGCCTC M.avium
1     ---------TTAACACATGCAAGTAGAACGGAAAGACCCC M.intracellulare
39    GCGGCGTGCTTAACACATGCAAGTCGAACGGAAAGGCCTC M.paratuberc.
2     GCGGCGTGCTTAACAAATGCAAGTCGAACGGAAAGGCCCC M.scrofulaceum
40    GCGGCGTGCTTAACACATGCAAGTCGAACGGAAAGGTCTC M.leprae
2     -CGGCGTGCTTAACACATGCAAGTCGAACGGAAAGGTCTC M.kansasii
2     GCGGCGTGCTTAACACATGCAAGTCGAACGGAAAGGTCTC M.gastri

                          90        100       110       120
42    T------------TCGGAGATACTCGAGTGGCGAACGGGT M.tuberculosis
181   T------------TCGGAGATACTCGAGTGGCGAACGGGT M.bovis
79    T------------TCGGAGGTACTCGAGTGGCGAACGGGT M.avium
32    T------------TCGGGG-TACTCGAGTGGCGAACGGGT M.intracellulare
79    T------------TCGGAGGTACTCGAGTGGCGAACGGGT M.paratuberc.
42    T------------TCGGGGGTACTCGAGTGGCGAACGGGT M.scrofulaceum
80    TAAAAAATCTTTTTTAGAGATACTCGAGTGGCGAACGGGT M.leprae
41    T------------TCGGAGACACTCGAGTGGCGAACGGGT M.kansasii
42    T------------TCGGAGACACTCGAGTGGCGAACGGGT M.gastri

                          130       140       150       160
70    GAGTAACACGTGGGTGATCTGCCCTGCACTTC-GGGATAA M.tuberculosis
209   GAGTAACACGTGGGTGATCTGCCCTGCACTTC-GGGATAA M.bovis
107   GAGTAACACGTGGGCAATCTGCCCTGCACTTC-GGGATAA M.avium
59    GAGTAACACGTGGGCAATCTGCCCTGCACTTC-GGGATAA M.intracellulare
107   GAGTAACACGTGGGCAATCTACCCTGCACTTC-GGGATAA M.paratuberc.
70    GAGTAACACGTGGGCAATCTGCCCTGCACTTC-GGGATAA M.scrofulaceum
120   GAGTAACACGTGGGTAATCTGCCCTGCACTTCAGGGATAA M.leprae
69    GAGTAACACGTGGGCAATCTGCCCTGCACACC-GGGATAA M.kansasii
70    GAGTAACACGTGGGCAATCTGCCCTGCACACC-GGGATAA M.gastri
```

Figure 2A

```
              170       180       190       200
109  GCCTGGGAAACTGGGTCTAATACCGGATAGGACCACGGGA M.tuberculosis
248  GCCTGGGAAACTGGGTCTAATACCGGATAGGACCACGGGA M.bovis
146  GCCTGGGAAACTGGGTCTAATACCGGATAGGACCTCAAGA M.avium
98   GCCTGGGAAACTGGGTCTAATACCGGATAGGACCTTTAGG M.intracellulare
146  GCCTGGGAAACTGGGTCTAATACCGGATAGGACCTCAAGA M.paratuberc.
109  GCCTGGGAAACTGGGTCTAATACCGGATAGGACCACTTGG M.scrofulaceum
160  GCTTGGGAAACTGGGTCTAATACCGGATAGGACTTCAAGG M.leprae
108  GCCTGGGAAACTGGGTCTAATACCGGATAGGACCACTTGG M.kansasii
109  GCCTGGGAAACTGGGTCTAATACCGGATAGGACCACTTGG M.gastri

              210       220       230       240
149  TGCATGTCTTGTGGTGGAAAGCGCTTTAGCGGTGTGGGAT M.tuberculosis
288  TGCATGTCTTGTGGTGGAAAGCGCTTTAGCGGTGTGGGAT M.bovis
186  CGCATGTCTTCTGGTGGAAAGC-TTTT-ACGGTGTGGGAT M.avium
138  CGCATGTCTTTAGGTGGAAAG--CTTTTGCGGTGTGGGAT M.intracellulare
186  CGCATGTCTTCTGGTGGAAAGC-TTTT-GCGGTGTAGAAT M.paratuberc.
149  CGCATGCCTTGTGGTGGAAAG--CTTTTGCGGTGTGGGAT M.scrofulaceum
200  CGCATGTCTTGTGGTGGAAAGC-TTTTTGCGGTGCAGGAT M.leprae
148  CGCATGCCTTGTGGTGGAAAG--CTTTTGCGGTGTGGGAT M.kansasii
149  CGCATGCCTTGTGGTGGAAAG--CTTTTGCGGTGTGGGAT M.gastri

              250       260       270       280
189  GAGCCCGCGGCCTATCAGCTTGTTGGTGGGGTGACGGCCT M.tuberculosis
328  GAGCCCGCGGCCTATCAGCTTGTTGGTGGGGTGACGGCCT M.bovis
224  GGGCCCGCGGCCTATCAGCTTGTTGGTGGGGTGACGGCCT M.avium
176  GGGCCCGCGGCCTATCAGCTTGTTGGTGGGGTGATGGCCT M.intracellulare
224  GGGCCCGCGGCCTATCAGCTTGTTGGTGGGGTGACGGCCT M.paratuberc.
187  GGGCCCGCGGCCTATCAGCTAGTTGGTGGGGTGATGGCCT M.scrofulaceum
239  GGGCCCGCGGCCTATCAGCTAATTAGTGGGGTAACGGCCT M.leprae
186  GGGCCCGCGGCCTATCAGCTTGTTGGTGGGGTGACGGCCT M.kansasii
187  GGGCCCGCGGCCTATCAGCTTGTTGGTGGGGTGACGGCCT M.gastri
```

Figure 2B

```
                450        460        470        480
389   AAACCTCTTTCACCATCGACGAAGGTCCGGGTTCTCTCGG M.tuberculosis
528   AAACCTCTTTCACCATCGACGAAGGTCCGGGTTCTCTCGG M.bovis
424   AAACCTCTTTCACCATCGACGAAGGTCCGGGTTTTCTCGG M.avium
376   AAACCTCTTTCACCATCGACGAAGGTCCGGGTTTTCTCGG M.intracellulare
424   AAACCTCTTTCACCATCGACGAAGGTCCGGGTTTTCTAGG M.paratuberc.
387   AAACCTCTTTCACCATCGACGAAGGCTCA---CTTTGTGG M.scrofulaceum
439   AAACCTCTTTCACCATCGACGAAGGTCTGGGAATTCTCGG M.leprae
386   AAACCTCTTTCACCATCGACGAAGGTCCGGGTTCTCTCGG M.kansasii
387   AAACCTCTTTCACCATCGACGAAGGTCCGGGTTCTCTCGG M.gastri
```

• • •

```
                1130       1140       1150       1160
1069  TCTCATGTTGCCAGCACGTAATGGTGGGGACTCGTGAGAG M.tuberculosis
1208  TCTCATGTTGCCAGCACGTAATGGTGGGGACTCGTGAGAG M.bovis
1104  TCTCATGTTGCCAGCGGGTAATGCCGGGGACTCGTGAGAG M.avium
1056  TCTCATGTTGCCAGCGGGTAATGCCGGGGACTCGTGAGAG M.intracellulare
1098  TCTCATGTTGCCAGCGGGTAATGCAGGGGACTCGTGAGAG M.paratuberc.
1064  TCTCATGTTGCCAGCGGGTAATGCCGGGGACTCGTGAGAG M.scrofulaceum
1119  TCTCATGTTGCCAGCACGTAATGGTGGGGACTCGTGAGAG M.leprae
1066  TCTCATGTTGCCAGCGGGTAATGCCGGGGACTCGTGAGAG M.kansasii
1067  TCTCATGTTGCCAGCGGGTAATGCCGGGGACTCGTGAGAG M.gastri
```

• • •

```
                1250       1260       1270       1280
1189  CAATGGCCGGTACAAAGGGCTGCGATGCCGCGAGGTTAAG M.tuberculosis
1328  CAATGGCCGGTACAAAGGGCTGCGATGCCGCGAGGTTAAG M.bovis
1224  CAATGGCCGGTACAAAGGGCTGCGATGCCGTAAGGTTAAG M.avium
1176  CAATGGCCGGTACAAAGGGCTGCGATGCCGCAAGGTTAAG M.intracellulare
1218  CAATGGCCGGTACAAAGGGCTGCGATGCCGTAAGGTTAAG M.paratuberc.
1184  CAATGGCCGGTACAAAGGGCTGCGATGCCGCAAGGTTAAG M.scrofulaceum
1239  CAATGGCCGGTACAAAGGGCTGCGATGCCGCAAGGTTAAG M.leprae
1186  CAATGGCCGGTACAAAGGGCTGCGATGCCGCGAGGTTAAG M.kansasii
1187  CAATGGCCGGTACAAAGGGCTGCGATGCCGCGAGGTTAAG M.gastri
```

Figure 2C

```
                1290      1300      1310      1320
1229 CGAATCCTTA-AAAGCCGGTCTCAGTTCGGATCGGGGTCT M.tuberculosis
1368 CGAATCCTTA-AAAGCCGGTCTCAGTTCGGATCGGGGTCT M.bovis
1264 CGAATCCTTTTAAAGCCGGACTCAGTTCGGATTGGGGTCT M.avium
1216 CGAATCCTTTTAAAGCCGGTCTCAGTTCGGATTGGGGTCT M.intracellulare
1258 CGAATCCTTTTAAAGCCGGACTCAGTTCGGATTGGGGTCT M.paratuberc.
1224 CGAATCCTTTTAAAGCCGGTCTCAGTTCGGATCGGGGTCT M.scrofulaceum
1279 CGAATCCTTTTAAAGCCGGTCTCAGTTCGGATCGGGGTCT M.leprae
1226 CGAATCCTTTTAAAGCCGGTCTCAGTTCGGATCGGGGTCT M.kansasii
1227 CGAATCCTTTTAAAGCCGGTCTCAGTTCGGATCGGGGTCT M.gastri

                1330      1340      1350      1360
1268 GCAACTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCA M.tuberculosis
1407 GCAACTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCA M.bovis
1304 GCAACTCGACCCCATGAAGTCGGAGTCGCTAGTAATCGCA M.avium
1256 GCAACTCGACCCCATGAAGTCGGAGTCGCTAGTAATCGCA M.intracellulare
1298 GCAACTAGACCCAATGAAGTCGGAGTCGCTAGTAATCGCA M.paratuberc.
1264 GCAACTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCA M.scrofulaceum
1319 GCAACTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCA M.leprae
1266 GCAACTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCA M.kansasii
1267 GCAACTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCA M.gastri
```

Figure 2D

```
                  90        100        110        120
168   TGCCCCTCCGGG----TGGAAAAGTAGGACACCGCCGAAC M.tuberculcsis
79    TGCCCCTCCGGGG---TGGAAAAGTAGGGCACCGCCGAAC M.bovis
81    TGCCCTCACCGGG---TGGAAAAGTAGGACACCGCCGAAC M.phlei
3599  TGCCCATTCGCG----TGGAAAAGTAGGACACTGCCGAAC M.leprae
5782  TACCCTT-CCGGG---TGGAAAAGTAGGACACCGCCGAAC M.smegmatis
```

Figure 3

```
              90        100       110       120
382   GGGAGCTGTCAACCGAGCATTGATCCGAGGATTTCCGAAT M.avium
382   GGGAGCTGTCAACCGAGCATTGATCCGAGGATTTCCGAAT M.paratuberc.
1053  GGGAGCTGTCAACCGAGCGTGGATCCGAGGATTTCCGAAT M.tuberculosis
467   GGGAGCTGTCAACCGAGCGTGGATCCGAGGATTTCCGAAT M.phlei
392   GGGAGCTGTCAACCGAGCGTGGATCCGAGGATTTCCGAAT M.leprae
167   GGGAGCTGTCAACCGAGCGTGGATCCGAGGATTTCCGAAT M.gastri
110   GGGAGCTGTCAACCGAGCGTGGATCCGAGGATTTCCGAAT M.kansasii

2548  GGGAGCTGTCAACCGAGCGTTGATCCGAGGATGTCCGAAT M.smegmatis
```

• • •

```
              170       180       190       200
462   GAATATATAGGGTGCG-GGAGGTAACGCGGGGAAGTGAAA M.avium
462   GAATATATAGGGTGCG-GGAGGTAACGCGGGGAAGTGAAA M.paratuberc.
1133  GAATATATAGGGTGCG-GGAGGGAACGCGGGGAAGTGAAA M.tuberculosis
547   GAATATATAGGCGTTG-GGGGGGAACGCGGGGAAGTGAAA M.phlei
472   GAATATATAGGGTTCG-GGAGGGAACGCGGGGAAGTGAAA M.leprae
247   GAATATATAGGGTGCG-GGAGGGAACGCGGGGAAGTGAAA M.gastri
190   GAATATATAGGGTGCG-GGAGGGAACGCGGGGAAGTGAAA M.kansasii

2628  GAATATATAGGCGTCT-GGGGGGAACGCGGCGAAGTGAAA M.smegmatis
```

Figure 4A

```
                   250       260       270       280
541   -GTCAGTAGTGGCGAGCGAAC-CGGAACA-GGCTAAACCG M.avium
541   -GTCAGTAGTGGCGAGCGAAC-CGGAACA-GGCTAAACCG M.paratuberc.
1212  -GCAAGTAGTGGCGAGCGAACGCGGAACA-GGCTAAACCG M.tuberculosis
626   -GTGAGTAGTGGCGAGCGAA-AGGGAGGATGGCTAAACCG M.phlei
551   -GCAAGTAGTGGCGAGCGAACGTGGAATATGGCTAAACCG M.leprae
326   -GTCAGTAGTGGCGAGCGAACGCGGAACATGGCTAAACCG M.gastri
269   -GTAAGTAGTGGCGAGCGAACGCGGAACATGGCTAAACCG M.kansasii

2706  CGTGAGTAGTGGCGAGCGAACACGGAGGATGGCTAAAC-G M.smegmatis

                   290       300       310       320
578   CATG-CATGGACAACCGGGTAGGGGTTGTGTGTGCGGGGT M.avium
578   CATG-CATGGACAACCGGGTAGGGGTTGTGTGTGCGGGGT M.paratuberc.
1250  CACG-CATGGGTAACCGGGTAGGGGTTGTGTGTGCGGGGT M.tuberculosis
664   CGTG-CATGTGATACCCGGTCGGGGTTGTGTGTGCGGTGT M.phlei
590   CACA-CATGTCTAACTAGGTAGGGGTTGTGTGTGCGGTGT M.leprae
365   CACG-CATGGGTGACCGGGTAGGGGTTGTGTGTGCGGGGT M.gastri
308   CACG-CATGGGTAACCGGGTAGGGGTTGTGTGTGCGGGGT M.kansasii

2745  TATGACATGTGATACCGGGTAGGGGTTGTGTGTGCGGGGT M.smegmatis
```

Figure 4B

```
                       330       340       350       360
617   TGTGGGATTGATATGTCTCAGCTCTACCTGGCTGAGG-GG M.avium
617   TGTGGGATTGATATGTCTCAGCTCTACCTGGCTGAGG-GG M.paratuberc.
1289  TGTGGGAG-GATATGTCTCAGCGCTACCCGGCTGAGA-GG M.tuberculosis
703   TGTGGGGCCTGTGTGTC-CATCGTCCGCCGGCGATGGCAG M.phlei
629   TGTGGGATTGGTATGTCTCAACTCTACCTGGTTGAGG-GG M.leprae
404   TGTGGGATCGATACGTCTCAGCTCTACCCGGCTGAGG-GG M.gastri
347   TGTGGGATCGATACGTCTCAGCTCTACCCGGCTGAGG-GG M.kansasii

2785  TGTGGGACCTATCTTTC-CGCCTCTACCTGGCTG-GAGGG M.smegmatis

                       370       380       390       400
656   TAGTCAGAAAGTGTCGTGGTTAGCGGAAGTGGCCTGGGAC M.avium
656   TAGTCAGAAAGTGTCGTGGTTAGCGGAAGTGGCCTGGGAC M.paratuberc.
1327  CAGTCAGAAAGTGTCGTGGTTAGCGGAAGTGGCCTGGGAT M.tuberculosis
742   TAGTGATAAAGCAGTGTGGTTAGGTGAAGTGGCCTGGGAT M.phlei
668   TAGTCAGAAAGTGCCGTGGTTAGCGGAAATGGCCTGGGAT M.leprae
443   CAGTCAGAAAGTGTCGTGGTTAACGGAAGTGGCCTGGGAT M.gastri
386   CAGTCAGAAAGTGTCGTGGTTAACGGAAGTGGCCTGGGAT M.kansasii

2823  CAGTGAGAAAATGTTGTGGTTAGCGGAAATGGCTTGGGAT M.smegmatis

                       410       420       430       440
696   GGCCCGCCGTAGACGGTGAGAGCCCGGTACGCGAAA-ACC M.avium
696   GGCCCGCCGTAGACGGTGAGAGCCCGGTACGCGAAA-ACC M.paratuberc.
1367  GGTCTGCCGTAGACGGTGAGAGCCCGGTACGCGAAA-ACC M.tuberculosis
782   GGTCTGCCGTAGTGGGTGAGAGCCCGTAAC-CGAAA-ACA M.phlei
708   GGCCTGCCGTAGACGGTGAGAGCCCAGTACGCGAAA-GCC M.leprae
483   GGTCTGCCGTAGACGGTGAGAGCCCGGTACGTGAAA-ACC M.gastri
426   GGTCTGCCGTAGACGGTGAGAGCCCGGTACGTGAAA-ACC M.kansasii

2863  GGCCTCCCGTAGACGGTGAGAGCCCGGTACGTGAAA-ACC M.smegmatis
```

Figure 4C

```
                450       460       470       480
735   CGGCACCTGCCTTATATCAACACCCGAGTAGCAGCGGGCC M.avium
735   CGGCACCTGCCTTATATCAACACCCGAGTAGCAGCGGGCC M.paratuberc.
1406  CGGCACCTGCCTAGTATCAATTCCCGAGTAGCAGCGGGCC M.tuberculosis
820   TGCTGCC-GCTGTCACAGG--TCCCGAGTAGCAGCGGGCC M.phlei
747   TGGCACCTGCCTTGTATCAATTCCCGAGTAGCAGCGGGCC M.leprae
522   CGGCACCTGCCTTGTATCAATTCCCGAGTAGCAGCGGGCC M.gastri
465   CGGCACCTGCCTTGTATCAATTCCCGAGTAGCAGCGGGCC M.kansasii

2902  CGACGTCTGTCTTGATGGTGTTCCCGAGTAGCAGCGGGCC M.smegmatis
```

• • •

```
                570       580       590       600
855   GAGGGAATGGTGAAAAGTACCCCGGCAGGG-AGTGAAATA M.avium
855   GAGGGAATGGTGAAAAGTACCCCGGGAGGG-AGTGAAATA M.paratuberc.
1526  GAGGGAATGGTGAAAAGTACCCCGGGAGGGGAGTGAAAGA M.tuberculosis
937   GAGGGAATCGTGAAAAGTACCCCGGGAGGG-AGTGAAAGA M.phlei
867   GAGGGAATGGTGAAAAGTACCCCGGGAGGGGAGTGAAATA M.leprae
642   GAGGGAATGGTGAAAAGTACCCCGGGAGGGGAGTGAAAGA M.gastri
585   GAGGGAATGGTGAAAAGTACCCCGGGAGGGGAGTGAAAGA M.kansasii

3022  GAGGGAATGGTGAAAAGTACCCCGGGAGGGGAGTGAAAGA M.smegmatis
```

Figure 4D

```
                610        620        630        640
894   GTACCTGAAACCGTGTGCCTACAATCCGTCAGAGCCTCCT M.avium
894   GTACCTGAAACCGTGTGCCTACAATCCGTCAGAGCCTCCT M.paratuberc.
1566  GTACCTGAAACCGTGTGCCTACAATCCGTCAGAGCCTCCT M.tuberculosis
976   GTACCTGAAACCGTGTGCCTACAATCCGTCAAAGCCCTCT M.phlei
907   GTACCTGAAACCGTGTGCCTACAATCCGTCAGAGCCTCTT M.leprae
682   GTACCTGAAACCGTGTGCCTACAATCCGTCAGAGCCCCTT M.gastri
625   GTACCTGAAACCGTGTGCCTACAATCCGTCAGAGCCCTTT M.kansasii

3062  GTACCTGAAACCGTGCGCTTACAATCCGTCAGAGCCCTCG M.smegmatis

                650        660        670        680
934   C---------------GTGGGGTGATGGCGTGCCTTTTGA M.avium
934   C---------------GTGGGGTGATGGCGTGCCTTTTGA M.paratuberc.
1606  TTTCCTCTCCGGAGGAGGGTGGTGATGGCGTGCCTTTTGA M.tuberculosis
1016  CTT-----------GTAGTGGGGTGATGGCGTGCCTTTTGA M.phlei
947   T---------------GTGGGGTGATGGCGTGCTTTTTGA M.leprae
722   T---------------GTGGGGTGATGGCGTGCCTTTTGA M.gastri
665   C---------------GTGGGGTGATGGCGTGCCTTTTGA M.kansasii

3102  ACGTGT----------GTGGGGTGATGGCGTGCCTTTTGA M.smegmatis

                690        700        710        720
959   AGAATGAGCCTGCGAGTCAGGGACACGTCGCGAGGTTAAC M.avium
959   AGAATGAGCCTGCGAGTCAGGGACACGTCGCGAGGTTAAC M.paratuberc.
1646  AGAATGAGCCTGCGAGTCAGGGACATGTCGCAAGGTTAAC M.tuberculosis
1046  AGAATGAGCCTGCGAGTCAGGGACATGTCGCGAGGTTAAC M.phlei
972   AGAATGAGCCTGCGAGTCAGGGACATGTCGCGAGGTTAAC M.leprae
747   AGAATGAGCCTGCGAGTCAGGGACATGTCGCGAGGTTAAC M.gastri
690   AGAATGAGCCTGCGAGTCAGGGACATGTCGCGAGGTTAAC M.kansasii

3132  AGAATGAGCCTGCGAGTCAGGGACATGTCGCGAGGTTAAC M.smegmatis
```

Figure 4E

```
                    770        780        790        800
1039 CGCATCCCCTTTGGG-----------GTGTAGTGGCGTGT M.avium
1039 CGCATCCCCTTTTGGG----------GTGTAGTGGCGTGT M.paratuberc.
1726 CGACCCACACGCGCATACGCGCGTGTGAATAGTGGCGTGT M.tuberculosis
1126 CGTATCCAACCTGTT----GCGGTTGGTGTAGTGGTGTGT M.phlei
1052 CGTATCACGTGTGAGCGT--------GTGTAGTGGCGTGT M.leprae
827  CGTATCACGCGTAAGCGT--------GTGTAGTGGCGTGT M.gastri
770  CGTATCGCGCGCGAGCGT--------GTGTAGTGGCGTGT M.kansasii

3212 CGTATCCACACAAGAGTGTGTG----GTGTAGTGGTGTGT M.smegmatis
```

• • •

```
                    1050       1060       1070       1080
1307 CAGCCAAACTCCGAATGCCG-TGGTG-TAAAAGCGTGGCA M.avium
1307 CAGCCAAACTCCGAATGCCG-TGGTG-TAAAAGCGTGGCA M.paratuberc.
2005 CAGCCAAACTCCGAATGCCG-TGGTG-TA-AAGCGTGGCA M.tuberculosis
1401 CAGCCAAACTCCGAATGCCGATAAG--TGAAAGTGTGGCA M.phlei
1323 CAGCCAAACTCCGAATGCCG-TGGTT-TAAAAGCGTGGCA M.leprae
1098 CAGCCAAACTCCGAATGCCG-TGGTG-TATA-GCGTGGCA M.gastri
1041 CAGCCAAACTCCGAATGCCG-TGGTG-TATA-GCGTGGCA M.kansasii

3486 CAGCCAAACTCCGAATGCCGGTAAGGCCAAGAGTGCGGAA M.smegmatis
```

• • •

```
                    1170       1180       1190       1200
1425 AGTGGAAAAGGATGTGTAGTCGCAGA-GACAACCAGGAGG M.avium
1425 AGTGGAAAAGGATGTGTAGTCGCAGA-GACAACCAGGAGG M.paratuberc.
2122 AGTGGGAAAGGATGTGCAGTCGCAAA-GACAACCAGGAGG M.tuberculosis
1519 AGTGGAAAAGGATGTGCAGTCGC-GAAGACAACCAGGAGG M.phlei
1441 AGTGGAAAAGGATGTGCAGTCGCAAA-GACAACCAGGAGG M.leprae
1215 AGTGGGAAAGGATGTGCAGTCGCAGA-GACAACCAGGAGG M.gastri
1158 AGTGGGAAAGGATGTGCAGTCGCAGA-GACAACCAGGAGG M.kansasii

3606 AGTGGAAAAGGATGTGAAGTCGCAGAAGAAAACCAGGAGG M.smegmatis
```

Figure 4F

```
                1250      1260      1270      1280
1504 CTCACTGGTCAAGTGATTATGCGCCGATAATGTAGCGGGG M.avium
1504 CTCACTGGTCAAGTGATTATGCGCCGATAATGTAGCGGGG M.paratuberc.
2201 CTCACTGGTCAAGTGATTGTGCGCCGATAATGTAGCGGGG M.tuberculosis
1598 CTCACTGGTCAAGTGATTGTGCGCTGATAATGTAGCGGGG M.phlei
1520 CTCACTGGTCAAGTGATTGTGCGCCGATAATGTAGCGGGG M.leprae
1294 CTCACTGGTCAAGTGATTGTGCGCCGATAATGTAGCGGGG M.gastri
1237 CTCACTGGTCAAGTGATTGTGCGCCGATAATGTAGCGGGG M.kansasii

3686 TTCACTGGTCAAGTGATTGTGCGCCGATATTGTGGCGGGG M.smegmatis

                1290      1300      1310      1320
1544 CTCAAGCACACCGCCGAAGCCGCGGCACATTCATCTT-TA M.avium
1544 CTCAAGCACACCGCCGAAGCCGCGGCACATTCATCTT-TA M.paratuberc.
2241 CTCAAGCACACCGCCGAAGCCGCGGCACATCCACCTTGT- M.tuberculosis
1638 CTCAAGCACACCGCCGAAGCCGCGGCA--ATCAGCCTTTG M.phlei
1560 CTCAAGCACACCGCCGAAGCCGCGGCACATTCACCTTCTA M.leprae
1334 CTCAAGCACACCGCCGAAGCCGCGACA-----ACCGC--A M.gastri
1277 CTCAAGCACACCGCCGAAGCCGCGACA-----ACCGC--A M.kansasii

3726 TTCAAGCACACCGCCGAAGCCGCGGAA--GCCAACGTTTG M.smegmatis

                1330      1340      1350      1360
1583 CGGTGGATGTGGGTAGGGGAGCGTCCCCCATTCAGCGAAG M.avium
1583 CGGTGGATGTGGGTAGGGGAGCGTCCCCCATTCAGCGAAG M.paratuberc.
2280 -GGTGGGTGTGGGTAGGGGAGCGTCCCTCATTCAGCGAAG M.tuberculosis
1676 TGGCTGGTGTGGGTAGGGGAGCGTCCTGCATCCGGTGAAG M.phlei
1600 GGGTGGATGTGGGTAGGGGAGCGTTCCTCATTCAGCGAAG M.leprae
1367 AGGT-----TGGGTAGGGGAGCGTCCCTCATTCAGCGAAG M.gastri
1310 AGGT-----TGGGTAGGGGAGCGTCCCTCATTCAGCGAAG M.kansasii

3764 TT-------TGGGTAGGGGAGCGTCCTG-ATCCGGTGAAG M.smegmatis
```

Figure 4G

```
                1370      1380      1390      1400
1623 CT-CCGGGTGACCGGTGGTGGAGGGTGGGGAGTGAGAAT M.avium
1623 CT-CCGGGTGATCGGTGGTGGAGGGTGGGGAGTGAGAAT M.paratuberc.
2319 CCACCGGGTGACCGGTGGTGGAGGGTGGGGAGTGAGAAT M.tuberculosis
1716 CCGCCGAGTGATCGGTGGTGGAGGGTGTGGGAGTGAGAAT M.phlei
1640 CCTCCGGGTAACCGGTGGTGGAGGGTGGGAAGTGAGAAT M.leprae
1402 CCGCCGGGTGACCGGTGGTGGAGGATGGGGAGTGAGAAT M.gastri
1345 CTGCCGGGTGACCGGTGGTGGAGGATGGGGAGTGAGAAT M.kansasii

3796 CCGCCGAGTATCGAGTGGTGGAGGGTGTGGGAGTGAGAAT M.smegmatis
```

• • •

```
                1530      1540      1550      1560
1781 CGATGGACAACGGGTTGATATTCCCGTACCCGTGTATGGG M.avium
1781 CGATGGACAACGGGTTGATATTCCCGTACCCGTGTATGGG M.paratuberc.
2479 CGATGGACAACGGGTTGATATTCCCGTACCCGTGTGTGGG M.tuberculosis
1875 CGATGGACAACGGGTTGATATTCCCGTACCCGTGTATGAG M.phlei
1800 CGATGGACAACGGGTTGATATTCCCGTACCCGTGTGTGTG M.leprae
1562 CGATGGACAACGGGTTGATATTCCCGTACCCGTGTGTGGG M.gastri
1505 CGATGGACAACGGGTTGATATTCCCGTACCCGTGTGTGGG M.kansasii

3956 CGATGGACAACGGGTTGATATTCCCGTACCCGTGTATGTG M.smegmatis

                1570      1580      1590      1600
1821 CGTCCCTGATGAATCA-GCGGTACTAACCACCCAAAACCG M.avium
1821 CGTCCCTGATGAATCA-GCGGTACTAACCACCCAAAACCG M.paratuberc.
2519 CGCCCGTGACGAATCA-GCGGTACTAACCACCCAAAACCG M.tuberculosis
1915 CGTCCCTGATGAATCTCATTCTGCTAACCACCCAAAACCT M.phlei
1840 CGCCCGTGATGAATCA-GCGGTACTCACCACCCAAAACCG M.leprae
1602 CGCCCGTGATGAATCA-GCGGTACTAACCACCCAAAACCG M.gastri
1545 CGCCCGTGATGAATCA-GCGGTACTAACCACCCAAAACCG M.kansasii

3996 CGTCCATGATGAATCA-GCGGTACTAACCATCCAAAACCA M.smegmatis
```

Figure 4H

```
               1610      1620      1630      1640
1860 GAT-CGACCAT-TCCCCTTCGGGGC-GTGGCGATT-CGG M.avium
1860 GAT-CGACCAT-TCCCCTTCGGGGC-GTGGCGATT-CGG M.paratuberc.
2558 GAT-CGATCAC-TCCCCTTCGGGG-TGTGGAGTTC-TGG M.tuberculosis
1955 GGC-CGATC--ATCC--TTCGGGG---GTGACGGTTG-GG M.phlei
1879 GAT-CGACCATATCCCCTTCGGGGCTATGGAGGTT-CGG M.leprae
1641 GAT-CGATCAC-TCCCCTTCGGGGA-GTGGAGGTC-TGG M.gastri
1584 GAT-CGATCAC-TCCCCTTCGGGGC-GTGGAGGTC-TGG M.kansasii

4035 ACCGTGACCGCACCT--TTCGGGG--TGTGGCGTTGGTGG M.smegmatis

               1650      1660      1670      1680
1896 GGCTGCGTGGGACCTTCGCTGGTAGTAGTCAAGCAATGGG M.avium
1896 GGCTGCGTGGGACCTTCGCTGGTAGTAGTCAAGCAATGGG M.paratuberc.
2594 GGCTGCGTGGGAACTTCGCTGGTAGTAGTCAAGCGAAGGG M.tuberculosis
1986 GGCTGCGTGGGACCCG-GTGGGTAGTAGTCAAGCGATGGG M.phlei
1917 GGCTGCGTGGGAACTTCGTTGGTAGTAGTCAAGCGATGGG M.leprae
1677 GGCTGCGTGGAGCCTTCGCTGGTAGTAGTCAAGCGATGGG M.gastri
1620 GGCTGCGTGGAGCCTTCGCTGGTAGTAGTCAAGCGATGGG M.kansasii

4071 GGCTGCATGGGACCTTCGTTGGTAGTAGTCAAGCGATGGG M.smegmatis
```

• • •

```
               1730      1740      1750      1760
1974 -CTGGGGCAAGCCCGTAG--AGAGCGATAGGCAAATCCGT M.avium
1974 -CTGGGGCAAGCCCGTAG--AGAGCGATAGGCAAATCCGT M.paratuberc.
2672 -CTGGGGCAAGCCGGTAGGGAGAGCGATAGGCAAATCCGT M.tuberculosis
2063 -CCGGGGTAAACCTGTAGGGCGAGTGATAGGCAAATCCGT M.phlei
1995 -CTGGAGCAAGCCCGTAGGGAGAGCGATAGGCAAATCCGT M.leprae
1755 -CTGGGGCAAGCCAGTAGGGAGAGCGATAGGCAAATCCGT M.gastri
1698 -CTGGGGCAAGCCAGTAGGGAGAGCGATAGGCAAATCCGT M.kansasii

4149 -CCGGCGTAAGCCTGTAGGGAGTCAGATAGGTAAATCCGT M.smegmatis
```

Figure 4I

```
             1850       1860       1870       1880
2089 GCGAGCACATACACGGCCCGTACCCCAAACCAACACAGGT M.avium
2089 GCGAGCACATACACTGCCCGTACCCCAAACCAACACAGGT M.paratuberc.
2789 GCGAGCACACACACGGCCCGTACCCCAAACCGACACAGGT M.tuberculosis
2179 GCAAGCGCATACACGGCCCGTACCCCAAACCAACACAGGT M.phlei
2112 GCGAGCATACATGCGGCCCGTACCCCAAACCGACACAGGT M.leprae
1872 GCGAGCACACACACGGCCCGTACCCCAAACCGACACAGG  M.gastri
1815 GCGAGCACACACACGGCCCGTACCCCAAACCGACACAGGT M.kansasii

4266 GCGAGGACATACACGGCCCGTACCCCAAACCAACACAGGT M.smegmatis
```

• • •

```
             1970       1980       1990       2000
2208 AGGGGGCCCGGAATACCGTGAACACCCTTGCGGTGGAGC M.avium
2208 AGGGGGCCCGGAATACCGTGAACACCCTTGCGGTGGAGC M.paratuberc.
2908 AGGGGGACCGGAATATCGTGAACACCCTTGCGGTGGAGC M.tuberculosis
2298 AGGGGGACCCACGTACCGTGAGGGCTCTTGCGGCGGCAGC M.phlei
2231 AGGGGGGCCGGAATATCGTGAACACCCTTGCGGTGGAGC M.leprae
1910                                         M.gastri
1934 AGGGGGACCGGAATACCGTGAACACCCTTGCGGTGGAGC M.kansasii

4385 AGGGGGACCCACATGGCGTGTAAGCCTTTACGGCCCAAGC M.smegmatis

             2010       2020       2030       2040
2248 GGGATTCGGCCGCAGAAACCAGTGGGTAGCGACT-GTTTA M.avium
2248 GGGATTCGGCCGCAGAAACCAGTGGGTAGCGACT-GTTTA M.paratuberc.
2948 GGGATCCGGTCGCAGAAACCAGTGAGGAGCGACT-GTTTA M.tuberculosis
2338 GGGGGTGGGTGGCACAAACCAGTGAGGAGCGACT-GTTTA M.phlei
2271 GGGATCCGGTCGCAGAGACCAGTGAGAAGCGACT-GTTTA M.leprae
1910                                          M.gastri
1974 GGGATTCGGTCGCAGAAACCAGTGAGAAGCGACTTGTTTA M.kansasii

4425 GTGAGTGGGTGGCACAAACCAGTGAGAAGCGACT-GTTTA M.smegmatis
```

Figure 4J

```
            2130      2140      2150      2160
2367 CCGTTAACCCGT--AAGGGTGAAGCGGAGAATTTAAGCCC M.avium
2367 CCGTTAACCCGT--AAGGGTGAAGCGGAGAATTTAAGCCC M.paratuberc.
3067 CCGTTAACCCGC--AAGGGTGAAGCGGAGAATTTAAGCCC M.tuberculosis
2457 CCGTTAACCCTTTCGGGGGTGAAGCGGAGAATTTAAGCCC M.phlei
2390 CTGTTAACCCGA--AAGGGTGAAGCGGAGAATTTAAGCCC M.leprae
1910                                          M.gastri
2094 CCGTTAACCCGC--AAGGGTGAAGCGGAGAATTTAAGCCC M.kansasii

4544 CCGTTAACCCCCTTGGGGGTGAAGCGGAGAATTTAAGCCC M.smegmatis
```

• • •

```
            2250      2260      2270      2280
2485 GTAACGACTTCCCAACTGTCTCAACCATAGACTCGGCGAA M.avium
2485 GTAACGACTTCCCAACTGTCTCAACCATAGACTCGGCGAA M.paratuberc.
3185 GTAACGACTTCTCAACTGTCTCAACCATAGACTCGGCGAA M.tuberculosis
2577 GTAACGACTTCTCAACTGTCTCAACCATAGACTCGGCGAA M.phlei
2508 GTAACGACTTCTCAACTGTCTCAACCATAGACTCGGCGAA M.leprae
1910                                          M.gastri
2212 GTAACGACTTCTCAACTGTCTCAACCATAGACTCGGCGAA M.kansasii

4663 GTAACGACTTCTCAACTGTCTCAAC-ATAGACTCGGCGAA M.smegmatis
```

Figure 4K

```
                                2370      2380      2390      2400
2605 GTTCGGTACGGTTTGTGTAGGATAGGTGGGAGACTTTGAA M.avium
2605 GTTCGGTACGGTTTGTGTAGGATAGGTGGGAGACTTTGAA M.paratuberc.
3305 GTTCGGTACGGTTTGTGTAGGATAGGTGGGAGACTGTGAA M.tuberculosis
2697 GCTCGATACGGTTTGTGTAGGATAGGTGGGAGACTGTGAA M.phlei
2628 GTTCGGTGCGGTTTGTGTAGGATAGGTGGGAGACTGTGAA M.leprae
1910                                          M.gastri
2332 GTTCGGTACGGTTTGTGTAGGATAGGTGGGAGACTGTGAA M.kansasii

4782 GCTCGATACGGTTTGTGTAGGATAGGTGGGAGACTGTGAA M.smegmatis

                                2410      2420      2430      2440
2645 GCACAGACGCCAGTTTGTGTGGAGTCGTTGTTGAAATACC M.avium
2645 GCACAGACGCCAGTTTGTGTGGAGTCGTTGTTGAAATACC M.paratuberc.
3345 ACCTCGACGCCAGTTGCGGCGGAGTCGTTGTTGAAATACC M.tuberculosis
2737 GCTCGGACGCCAGTTCGGGTGGAGTCGTTGTTGAAATACC M.phlei
2668 ACTTCGACGCTAGTTGCGGTGGAGTCGTTGTTGAAATACC M.leprae
1910                                          M.gastri
2372 ACCTCAACGCCAGTTGCGGTGGAGTCGTTGTTGAAATACC M.kansasii

4822 GCTCACACGCCAGTGTGGGTGGAGTCGTTGTTGAAATACC M.smegmatis

                                2450      2460      2470      2480
2685 ACTCTGATCGTATTGGACACCTAACGTCGAACCCT-TATC M.avium
2685 ACTCTGATCGTATTGGACACCTAACGTCGAACCCT-TATC M.paratuberc.
3385 ACTCTGATCGTATTGGGCATCTAACCTCGAACCCTGAATC M.tuberculosis
2777 ACTCTGATCGTATTGGGCCTCTAACCTCGGACCGTGGATC M.phlei
2708 ACTCTGATTGTATTGAACATCTAACCTCGAACCGTATATC M.leprae
1910                                          M.gastri
2412 ACTCTGATCGTATTGGACACCTAACGTCGAACCCTGAATC M.kansasii

4862 ACTCTGATCGTATTGGGCCTCTAACCTCGGACCGTATATC M.smegmatis
```

Figure 4L

```
             2690      2700      2710      2720
2924 GGTGTCACTCAACGGATAAAAGGTACCCCGGGGATAACGG M.avium
2924 GGTGTCACTCAACGGATAAAAGGTACCCCGGGGATAACAG M.paratuberc.
3625 GGTGTCGCTCAACGGATAAAAGGTACCCCGGGGATAACAG M.tuberculosis
3017 GGTGTCGCTCAACGGATAAAAGGTACCCCGGGGATAACAG M.phlei
2948 GGTGTCGCTCAACGGATAAAAGGTACCCCGGGGATAACAG M.leprae
1910                                          M.gastri
2652 GGTGTCGCTCAACGGATAAAAGGTACCCCGGGGATAACAG M.kansasii

5102 GGTGTCGCTCAACGGATAAAAGGTACCCCGGGGATAACAG M.smegmatis
```

• • •

```
             2810      2820      2830      2840
3044 GGTCCCAAAGGTTGGGCTGTTCGCCC-ATTAAAGCGGCAC M.avium
3044 GGTCCCAAGGGTTGGGCTGTTCGCCC-ATTAAAGCGGCAC M.paratuberc.
3745 GGTCCCAAGGGTTGGGCTGTTCGCCC-ATTAAAGCGGCAC M.tuberculosis
3137 GGTCCCAAGGGTTGGGCTGTTCGCCC-ATTAAAGCGGCAC M.phlei
3068 GGTCCCAAGGGTTGGGCTGTTCGCCC-ATTAAAGCGGCAC M.leprae
1910                                          M.gastri
2772 GGTCCCAAGGGTTGGGCTGTTCGCCC-ATTAAAGCGGCAC M.kansasii

5222 GGTCCCAAGGGTTGGGCTGTTCGCCCCATTAAAGCGGCAC M.smegmatis
```

Figure 4M

```
                3050      3060      3070      3080
3283 CAAGATCAGGTTTC-TCACCCTTTTAGAGGGATAAGGCCC M.avium
3283 CAAGATCAGGTTTC-TCACCCTTTTAGAGGGATAAGGCCC M.paratuberc.
3984 CAAGATCAGGTTTC-TCACCCACTTGGTGGGATAAGGCCC M.tuberculosis
3376 CAAGACCAGGCTTC-TCACCCTCTAGGAGGGATAAGGCCC M.phlei
3307 CAA                                      M.leprae
1910                                          M.gastri
3011 CAAGATCAGGCTTC-TCACCCACTTGGTGGGATAAGGCCC M.kansasii

5462 CAAGACCAGGCTTC-TCACCCTCTAGGAGGGATAAGGCCC M.smegmatis

                3090      3100      3110      3120
3322 CCCGC-AGACCACGGGATTGATAGGCCAGACCTGGAAGCT M.avium
3322 CCCGC-AGATCACGGGATTGATAGGCCAGACCTGGAAGCT M.paratuberc.
4023 CCCGC-AGAACACGGGTTCAATAGGTCAGACCTGGAAGCT M.tuberculosis
3415 CCCGC-AGACCACGGGATCGATAGACCAGACCTGCACGCA M.phlei
3309                                          M.leprae
1910                                          M.gastri
3050 CCCGC-AGAACACGGGTTCGATAGGCCAGACCTGGAAGCT M.kansasii

5501 CCCGC-AGACCACGGGATTGATAGACCAGACCTGGAAGCC M.smegmatis
```

Figure 4N

```
                 130       140       150       160
107   GAGTAACACGTGGGCAATCTGCCCTGCACTTC-GGGATAA M.avium
59    GAGTAACACGTGGGCAATCTGCCCTGCACTTC-GGGATAA M.intracellulare
107   GAGTAACACGTGGGCAATCTACCCTGCACTTC-GGGATAA M.paratuberc.
70    GAGTAACACGTGGGCAATCTGCCCTGCACTTC-GGGATAA M.scrofulaceum
70    GAGTAACACGTGGGTGATCTGCCCTGCACTTC-GGGATAA M.tuberculosis
209   GAGTAACACGTGGGTGATCTGCCCTGCACTTC-GGGATAA M.bovis
120   GAGTAACACGTGGGTAATCTGCCCTGCACTTCAGGGATAA M.leprae
69    GAGTAACACGTGGGCAATCTGCCCTGCACACC-GGGATAA M.kansasii
70    GAGTAACACGTGGGCAATCTGCCCTGCACACC-GGGATAA M.gastri
```

• • •

```
                 450       460       470       480
424   AAACCTCTTTCACCATCGACGAAGGTCCGGGTTTTCTCGG M.avium
376   AAACCTCTTTCACCATCGACGAAGGTCCGGGTTTTCTCGG M.intracellulare
424   AAACCTCTTTCACCATCGACGAAGGTCCGGGTTTTCTAGG M.paratuberc.
387   AAACCTCTTTCACCATCGACGAAGGCTCA---CTTTGTGG M.scrofulaceum
389   AAACCTCTTTCACCATCGACGAAGGTCCGGGTTCTCTCGG M.tuberculosis
528   AAACCTCTTTCACCATCGACGAAGGTCCGGGTTCTCTCGG M.bovis
439   AAACCTCTTTCACCATCGACGAAGGTCTGGGAATTCTCGG M.leprae
386   AAACCTCTTTCACCATCGACGAAGGTCCGGGTTCTCTCGG M.kansasii
387   AAACCTCTTTCACCATCGACGAAGGTCCGGGTTCTCTCGG M.gastri
```

• • •

```
                 1130      1140      1150      1160
1104  TCTCATGTTGCCAGCGGGTAATGCCGGGGACTCGTGAGAG M.avium
1056  TCTCATGTTGCCAGCGGGTAATGCCGGGGACTCGTGAGAG M.intracellulare
1098  TCTCATGTTGCCAGCGGGTAATGCAGGGGACTCGTGAGAG M.paratuberc.
1064  TCTCATGTTGCCAGCGGGTAATGCCGGGGACTCGTGAGAG M.scrofulaceum
1069  TCTCATGTTGCCAGCACGTAATGGTGGGGACTCGTGAGAG M.tuberculosis
1208  TCTCATGTTGCCAGCACGTAATGGTGGGGACTCGTGAGAG M.bovis
1119  TCTCATGTTGCCAGCACGTAATGGTGGGGACTCGTGAGAG M.leprae
1066  TCTCATGTTGCCAGCGGGTAATGCCGGGGACTCGTGAGAG M.kansasii
1067  TCTCATGTTGCCAGCGGGTAATGCCGGGGACTCGTGAGAG M.gastri
```

Figure 5A

```
                 1290      1300      1310      1320
1264 CGAATCCTTTTAAAGCCGGACTCAGTTCGGATTGGGGTCT M.avium
1216 CGAATCCTTTTAAAGCCGGTCTCAGTTCGGATTGGGGTCT M.intracellulare
1258 CGAATCCTTTTAAAGCCGGACTCAGTTCGGATTGGGGTCT M.paratuberc.
1224 CGAATCCTTTTAAAGCCGGTCTCAGTTCGGATCGGGGTCT M.scrofulaceum
1229 CGAATCCTTA-AAAGCCGGTCTCAGTTCGGATCGGGGTCT M.tuberculosis
1368 CGAATCCTTA-AAAGCCGGTCTCAGTTCGGATCGGGGTCT M.bovis
1279 CGAATCCTTTTAAAGCCGGTCTCAGTTCGGATCGGGGTCT M.leprae
1226 CGAATCCTTTTAAAGCCGGTCTCAGTTCGGATCGGGGTCT M.kansasii
1227 CGAATCCTTTTAAAGCCGGTCTCAGTTCGGATCGGGGTCT M.gastri
```

```
                 1330      1340      1350      1360
1304 GCAACTCGACCCCATGAAGTCGGAGTCGCTAGTAATCGCA M.avium
1256 GCAACTCGACCCCATGAAGTCGGAGTCGCTAGTAATCGCA M.intracellulare
1298 GCAACTAGACCCAATGAAGTCGGAGTCGCTAGTAATCGCA M.paratuberc.
1264 GCAACTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCA M.scrofulaceum
1268 GCAACTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCA M.tuberculosis
1407 GCAACTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCA M.bovis
1319 GCAACTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCA M.leprae
1266 GCAACTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCA M.kansasii
1267 GCAACTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCA M.gastri
```

Figure 5B